The Normality Assumption in Coordination Games with Flexible Information Acquisition

Rigos, Alexandros

2018

Document Version:
Other version

Link to publication

Citation for published version (APA):
Rigos, A. (2018). *The Normality Assumption in Coordination Games with Flexible Information Acquisition*. (Working Papers; No. 2018:30).

Total number of authors:
1

Working Paper 2018:30

Department of Economics
School of Economics and Management

The Normality Assumption in Coordination Games with Flexible Information Acquisition

Alexandros Rigos

November 2018
Revised: March 2022

The Normality Assumption in Coordination Games with Flexible Information Acquisition

Alexandros Rigos*

March 2022

Abstract

Many economic models assume that random variables follow normal (Gaussian) distributions. Yet, real-world variables may be non-normally distributed. How sensitive are these models' predictions to distribution misspecifications? This paper addresses the question in the context of linear-quadratic beauty contests played by rationally inattentive players. It breaks with the assumption that the (common prior) distribution of the fundamental be Gaussian and provides a characterization of the class of equilibria in continuous strategies. The characterization is used to show that small departures from normality can lead to distributions of the equilibrium average action that are qualitatively different from those of Gaussian models. Numerical results show that the rate at which an analyst's errors in determining the fundamental's distribution are amplified in her prediction is higher when the true prior is non-Gaussian than when it is an equally-misspecified Gaussian.

Keywords: Coordination games; Beauty contest; Flexible information acquisition; Rational inattention; Error amplification; Misspecified priors
JEL classification: C72, D83

---

*Institute for Futures Studies, Stockholm and Department of Economics, Lund University. Email: alexandros.rigos@gmail.com. Earlier versions of this paper have been circulated as "Flexible Information Acquisition in Large Coordination Games" and "A Beauty Contest with Flexible Information Acquisition." I thank Martin Kaae Jensen and Chris Wallace for their invaluable guidance and advice as well as the Editor, Xavier Vives, and two anonymous referees for constructive comments that pushed the paper in new directions. I also thank Mogens Fosgerau, Jens Gudmundsson, Erik Mohlin, Heinrich Nax, André Stenzel, Petra Thiemann, Jörgen Weibull, Christoph Carnehl, Ming Yang, and participants in numerous seminars, conferences, and workshops for useful comments and discussions. I am grateful to Erik Mohlin and Handelsbankens Forskningsstiftelser (grant #P2019-0204) for funding. All errors are mine.

1 Introduction

Many economic models assume that random variables follow normal (Gaussian) distributions. This assumption can be well-founded (because of the central limit theorem, for example), but it is often made because it makes the model tractable. This paper studies how sensitive predictions are to the normality assumption.

The paper focuses on beauty-contest games, a central class of games for applications of information economics in macroeconomics and finance (Angeletos and Pavan 2007; Morris and Shin 2002; Vives 2008). A payoff-relevant state (the fundamental) follows a commonly known prior over the real line. Each of many players takes an action (a real number) and loses payoff according to a weighted average of (a) the squared distance between her action and the realization of the fundamental and (b) the squared distance between her action and the population-wide average action. This paper studies how sensitive the model's predictions are to relaxing the widespread assumption that the payoff-relevant state and players' information (signals) be normally distributed.

A possible approach to the robustness question would be to characterize *all* equilibria that can arise across *all* possible joint distributions over states and signals, an approach in the spirit of Bergemann and Morris (2016). One could be worried that this approach, although compelling, leaves too much room for potential negative results. To put some discipline to the set of joint distributions, this paper assumes that players endogenously choose what information to acquire. Before choosing her action, each player can acquire information flexibly (à la Yang 2015). She can arbitrarily correlate her signal to the fundamental, while paying a cost that is linear in the Shannon mutual information between the two variables.

In this context, the choice of mutual information for the cost of information is natural. First, it is a widespread functional form in the literature on rational inattention (Sims 2003). Second, and more importantly, it enables the embedding of the standard Gaussian case into the analysis. When the fundamental is normally distributed, one can easily—by guessing and verifying—find equilibria whereby the fundamental and players' actions are jointly normally distributed and the average action is a linear function of the fundamental.

The paper's main technical contribution is a characterization of the class of equilibria in which players' actions follow continuous distributions. The characterization implies that the highly tractable equilibria in which the population-wide average action is a linear function of the funamental exist only if the fundamental is normally distributed. Based on this characterization, the paper proposes a novel method which allows the study of non-Gaussian priors and which is used to study the robustness of equilibrium predictions to misspecifications of the prior. The central finding of the paper is a negative result. Even small departures from a Gaussian prior can lead to distributions of equilibrium actions that differ significantly from the ones obtained for normal priors. The paper proceeds to quantify the sensitivity of the model's predictions to misspecifications of the fundamental's distribution.

The formal model is laid out in Section 2. Beauty contests and, more generally, linear-quadratic models lend themselves to the study of information acquisition in strategic environments (e.g. Dewan and Myatt 2008; Hellwig and Veldkamp 2009; Myatt and Wallace 2012).[1] In particular, they have been standard settings for models with rationally inattentive (à la Sims 2003) players (e.g. Hellwig, Kohls, and Veldkamp 2012; Maćkowiak and Wiederholt 2009; Hébert and La'O 2020). The critical assumption that makes many of these models tractable is that of a Gaussian prior.

Establishing the literature benchmark, the analysis begins in Section 3 with the case of Gaussian priors and jointly-normal strategies. As in Hellwig and Veldkamp (2009), players' motives for coordination translate into strategic complementarities in information acquisition. Importantly, the mutual-information cost function can make these complementarities strong enough for multiple equilibria to arise (see also Myatt and Wallace 2012; Hébert and La'O 2020).

Next, Section 4 parts with the assumption that the prior be Gaussian and finds a necessary and sufficient condition for a player to have a continuous best response to a well-behaved strategy profile of her opponents. The best response is such that after receiving any signal, the player's posterior belief about her optimal (or, *best*) action is a Gaussian centered at the action that the signal prescribes her to take.[2]

[1]A closely related literature considers information aggregation and acquisition in markets that consist of agents with linear-quadratic-loss or constant-absolute-risk-aversion (CARA) utility functions and where fundamental values and agents' signals are normally distributed (e.g. Grossman and Stiglitz 1980; Hellwig 1980). A succinct and informative exposition of these models can be found in Vives (2008).

[2]This echoes Matějka and Sims (2010), if one uses the best action as the tracked state (rather than

The characterization of equilibria in continuous strategies (Proposition 6, the paper's main technical result) is reached by demanding that the best-action distribution to which the players are best responding be equal to the one that is generated from their strategies, i.e., a fixed point. From the characterization, it follows that the model has affine closed-form solutions only under the normal prior assumption, which hints that some results derived for normal priors may not extend to other distributions. Even though equilibria for a given prior cannot be computed directly, the characterization result allows one to generate pairs of prior and equilibrium distributions. Based on this, Section 4.3 proposes a method to study equilibria under non-Gaussian priors.

The main results of the paper are laid out in Section 5, which studies the sensitivity of predictions to misspecifications of the prior. First, examples generated with the proposed method demonstrate that even when prior distributions are very close to the normal, the distributions of the equilibrium population-wide average action can be *qualitatively* different from those derived for a Gaussian prior. Second, numerical methods *quantify* how robust the predictions of the Gaussian model are to misspecifications of the prior. This is done by comparing the error that results in the prediction of the equilibrium average-action distribution (measured by the L^1 distance) to the error in the prior that caused it. The exercise shows that errors in the prior can be highly amplified in the model's predictions (by more than twenty-fold in some cases; Section 5.3). Moreover, using a misspecified Gaussian prior leads to larger error amplification rates when the true prior is non-Gaussian than when the true prior is an equally-misspecified Gaussian (Section 5.4). These error amplifications increase as coordination motives get stronger and as information gets costlier.

Prior misspecifications lead to larger prediction errors as information costs increase and as coordination motives get stronger. The intuition is the following. With costless information, players can perfectly observe—and, subsequently, successfully coordinate their actions on—any realization of the fundamental. As her information unit costs increase, though, a rationally inattentive player acquires less information. This leads to her paying attention to fewer features of the tracked variable's distribution and, consequently, to the distribution of her action getting more concentrated. Moreover,—because of the coordination motives present—as her opponents concentrate their ac-

---

the fundamental).

tions around some value, she optimally also takes actions close to that value more frequently. As a result, if the normal prior assumption is slightly wrong, the predicted equilibrium does not take account of the chain of "explosive" reasoning agents go through to pick their actions. So, even a small change in the distribution of the fundamental generates a chain reaction in complementary actions leading to an equilibrium relatively far from the one predicted on the basis of the Gaussian model.

Studying beauty contests and other economic models beyond the normality assumption can yield qualitatively novel results. For example, small departures from the normal prior in beauty contests can lead to highly skewed distributions of equilibrium actions. This means that if some endogenous variable (e.g. the average price) is observed to be highly skewed, this skewness can potentially be explained as the result of small non-Gaussian "distortions" of the prior, amplified by market participants' coordination motives. The normal prior assumption mutes by construction this type of new insights. The paper's results demonstrate that even though the normality assumption can make models highly tractable, their predictions are not always robust to distribution misspecifications. Analysts should therefore exercise caution when making the simplifying normality assumption.

The paper is closely related to Myatt and Wallace (2012), who study economic beauty contests with endogenous information acquisition, Yang (2015), who introduced flexible information acquisition technology, and Jung et al. (2019), who study how small changes in priors affect the optimal strategies of rationally inattentive agents. Another related study is Denti (2020), who allows players to design signals that can be correlated even after conditioning on the fundamental, while paying a cost that is increasing in the Blackwell order (Blackwell 1951, 1953). In this sense, his agents use an *unrestricted* information acquisition technology. The information structure used by Denti (2020) is, therefore, richer than the one used herein, and leads to Bayes correlated equilibria (Bergemann and Morris 2016). More recently, Hébert and La'O (2020) use a similar technology to study efficiency and non-fundamental volatility under different information costs. A detailed discussion of how results in the existing literature relate those presented here is postponed to Section 6.

2 Model

Consider a large population of ex-ante identical expected-payoff-maximizing players ($i \in [0,1]$). Players are Bayesian and have a common prior $P_\theta \in \Delta(\mathbb{R})$ about a payoff-relevant state of the world $\theta \in \mathbb{R}$, the *fundamental*.[3] The prior has full support over $\mathbb{R}$, well-defined mean $\bar{\theta}$ (without loss of generality $\bar{\theta} = 0$) and variance σ^2, and a probability density function (PDF) p, which is an analytic function. Each player i takes an action $a_i \in \mathbb{R}$, while the population-wide average action is denoted by $\bar{a} = \int_0^1 a_i \, \mathrm{d}i$.

Player objectives

The players have coordination motives with strength $\gamma \in [0,1)$ and fundamental motives with strength $1-\gamma$. They receive payoffs according to

$$-(1-\gamma)(a_i-\theta)^2 - \gamma(a_i-\bar{a})^2. \tag{1}$$

If a player knows θ and $\bar{a}$, she maximizes her payoff by playing the *best action*

$$b := (1-\gamma)\theta + \gamma\bar{a}. \tag{2}$$

Even when information is incomplete, the best action is very important to the players: it is the random variable that each of them is trying to match. To see this, notice that the objective (1) can be rewritten as $-(a_i-b)^2-\gamma(1-\gamma)(\bar{a}-\theta)^2$. Since player i maximizes expected payoff, she optimally chooses her action a_i so as to maximize

$$U_i := \mathbb{E}_i\left[-(a_i-b)^2\right]. \tag{3}$$

It follows that player i's optimal action is her expected best action, given her belief.

Observation 1 (Ex-post optimality). *A player i maximizes her payoff by playing $a_i = \mathbb{E}_i[b]$, where the expectation is taken according to her posterior belief.*

Information acquisition

Before choosing her action, and simultaneously with her opponents, each player i can acquire information about the fundamental. This information comes at a cost which

[3] Throughout the paper, $\Delta(X)$ denotes the set of probability measures over set X, while P_x denotes the distribution of variable x.

is linear in (Shannon) mutual information between the fundamental θ and her action a_i. Formally, an information acquisition strategy (or, simply, a *strategy*) is a family of conditional probability measures $P_{a_i|\theta}$ that give the distribution over actions (i.e. mixed strategy) conditional on the realization θ.[4] When the various $P_{a_i|\theta}$ admit densities, these are denoted by $r_i(\cdot|\theta)$ and information costs are given by

$$C_i := \mu \, I(r_i; p) = \mu \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(\theta) r_i(a_i|\theta) \log \frac{r_i(a_i|\theta)}{R_i(a_i)} \, \mathrm{d}a_i \, \mathrm{d}\theta \right), \tag{4}$$

where $R_i(a_i) = \int_{-\infty}^{+\infty} r_i(a_i|\theta) \, \mathrm{d}\theta$ is the marginal density of action a_i and $\mu \geq 0$ is a parameter which denotes the unit cost of information.[5]

A tuple (p, γ, μ) defines a *game,* which will occasionally be referred to as an *economy.*

Equilibrium

Given a strategy profile $r = (r_i)_{i \in [0,1]}$, the average action at a state θ is given by

$$\bar{a}(\theta) = \int_0^1 \int_{-\infty}^{+\infty} a_j r_j(a_j|\theta) \, \mathrm{d}a_j \, \mathrm{d}j,$$

which is assumed to be well-defined for all θ and (Lebesgue-)measurable.[6] Since a single player's choice of action cannot affect the population's average action, the function $\bar{a}(\cdot)$ is the object to which each player is best responding. Equivalently (for a fixed γ), player i is best responding to the best action function

$$b(\theta) := (1 - \gamma)\theta + \gamma \bar{a}(\theta).$$

It is now possible to define a notion of equilibrium. In an equilibrium, each player's strategy r_i should be optimal given $\bar{a}(\cdot)$ and the average action at any given state should be the aggregate of all players' actions, according to their strategies.

[4] More generally, information acquisition strategies map states of the world θ to distributions over *messages*. Standard arguments, though, establish that each message should just prescribe player i which action to take (see Online Appendix).

[5] As most of the paper's analysis focuses on strategies with densities, some definitions are stated for such strategies, so as to ease exposition. It should be understood, though, that the definitions extend to more general probability measures.

[6] Indeed, in all equilibria considered in the paper, the average action is well-defined and measurable.

Definition 1. *A strategy profile r^* is an equilibrium of the game (p,γ,μ) if*

$$r_i^* \in \underset{r_i \in (\Delta(\mathbb{R}))^{\mathbb{R}}}{\arg\max} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(-(1-\gamma)(a_i-\theta)^2 - \gamma(a_i - \bar{a}(\theta)^2)\right) r_i(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \mu\, I(r_i;p)$$

for all players i and the average action conditional on θ is given by

$$\bar{a}(\theta) = \int_0^1 \int_{-\infty}^{+\infty} a_j r_j^*(a_j|\theta)\,\mathrm{d}a_j\,\mathrm{d}j.$$

The main equilibrium object of interest is the distribution $P_{\bar{a}}$ of the average action, which summarizes the population's equilibrium aggregate behavior (and is likely to be observable in applications). If $\bar{a}(\cdot)$ is strictly increasing (which will be the case in all equilibria considered), then $P_{\bar{a}}$ has a density, denoted by h. This PDF can be calculated from the change of variables formula:

$$h(\cdot) = p\left(\bar{a}^{-1}(\cdot)\right)\left(\bar{a}^{-1}\right)'(\cdot). \tag{5}$$

With slight abuse, the term "h is an equilibrium of (p,γ,μ)" will be used occasionally.

2.1 Error amplification factor

A central aim of the paper is to study the sensitivity of equilibrium predictions to misspecifications of the prior p. This section introduces the measure that will be employed in Section 5 to quantify this sensitivity.

Consider an analyst who tries to derive predictions about an economy (p,γ,μ) of the type described above and who would make prediction h (an equilibrium of (p,γ,μ)) if she knew all three game parameters. Assume, now, that although the analyst knows the coordination motive γ and information cost μ, she does not know the distribution p of the fundamental (e.g., because she has observed only a finite amount of realizations of θ). Instead, she works under the (incorrect) assumption that the fundamental is distributed according to $\tilde{p}$ and predicts that the resulting average action distribution is $\tilde{h}$, an equilibrium of $(\tilde{p},\gamma,\mu)$. How wrong will her prediction $\tilde{h}$ be because of her using an incorrect prior $\tilde{p}$ in her calculations? Or, to put it differently: At what rate will her errors about the prior be translated into errors of her predictions?

In order to answer such questions, the notion of the *error amplification factor A* is used. It compares the error made in the predicted average-action distribution to the error made in the measurement of the prior distribution, using the L^1 norm.[7]

$$A = \frac{\left\|h - \tilde{h}\right\|_1}{\|p - \tilde{p}\|_1}$$

The higher the value of A, the more sensitive the model's predictions are to misspecifications of the prior.

Note that A depends on both the priors $p, \tilde{p}$ and the predictions $h, \tilde{h}$ that the analyst would make in case she knew the respective prior. If one were to define an amplification factor solely based on the priors $p, \tilde{p}$ instead, identification issues could arise when p or $\tilde{p}$ have multiple equilibria. In this light, any A value calculated for a pair of priors $p, \tilde{p}$ and the respective predictions $h, \tilde{h}$ should be interpreted as a lower bound of the amplification factor that can be reached under the prior pair $p, \tilde{p}$. That is, there might be another pair of predictions $h', \tilde{h}'$ that are equilibria of p and $\tilde{p}$, respectively, and for which the error amplification is even higher than the one calculated for $h, \tilde{h}$.

3 Equilibrium under a Gaussian prior

Before addressing the general problem in Section 4, a benchmark case that is popular in the literature is briefly presented. In this benchmark, the prior distribution is normal ($\theta \sim N(0, \sigma^2)$) and each player follows a strategy in which her expected action is linear in θ.[8] A formal definition of such strategies follows. In the definition, the degenerate $N(x, 0)$ is identified with the distribution that assigns all probability mass to x.

Definition 2 (Linear-Gaussian strategy)**.** *A strategy r_i of a player i will be called* linear-Gaussian *if by following r_i, player i's action conditional on the fundamental θ follows a Gaussian distribution with a mean that is linear in θ:*

$$a_i|\theta \sim N\left(\lambda_i \theta, \sigma^2_{a_i}\right) \text{ for some } \lambda_i \in [0,1] \text{ and some } \sigma^2_{a_i} \geq 0.$$

[7]The reader is reminded that $\|f\|_1 = \int_{-\infty}^{+\infty} |f(x)|\,dx$. Notice that since the total variation distance between two probability measures P and Q is equal to $\frac{1}{2}\|P - Q\|_1$, the measure A used here would be the same if total variation distance was used instead of the L^1 distance.

[8]In the literature (e.g. Myatt and Wallace 2012), the strategies described here are usually reached through the agents' updating of their (improper uniform) prior after observing a public signal and a private signal whose precision they can determine endogenously.

One, then, guesses and verifies that an equilibrium in linear-Gaussian strategies exists and proceeds with analyzing its properties.

Best response Indeed, if a player i's opponents follow linear-Gaussian strategies, then the average action to which she best responds is linear in θ (i.e., $\bar{a}(\theta) = \bar{\lambda}\theta$) with slope $\bar{\lambda} := \int_0^1 \lambda_j \,\mathrm{d}j \in [0,1]$. It follows that the best action is also linear (i.e., $b(\theta) = \kappa\theta$) with slope $\kappa = \big((1-\gamma)+\gamma\bar{\lambda}\big) \in [0,1]$. Under a Gaussian prior, player i's unique best response to such a profile is also linear-Gaussian (see, e.g., Sims 2003; Jung et al. 2019). Its slope is given by (see Appendix A.1 for all calculations)

$$\Lambda(\bar{\lambda}) = \begin{cases} \big((1-\gamma)+\gamma\bar{\lambda}\big) - \frac{\mu}{2\big((1-\gamma)+\gamma\bar{\lambda}\big)\sigma^2} & \text{if } \mu < 2\big((1-\gamma)+\gamma\bar{\lambda}\big)^2\sigma^2 \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

As her opponents' average action slope $\bar{\lambda}$ increases, player i increases her own slope λ_i in her best response. This happens for two reasons. First, a higher $\bar{\lambda}$ implies that the best action b—the variable that player i is trying to track—is more correlated to the fundamental. So, it is in her best interest to increase the correlation of her action with the fundamental too. Second, while player i acquires information about the fundamental θ, she uses this information to match the best action $b = \kappa\theta$. As her opponents' average action slope increases, b gets more correlated with θ and player i can use the information she acquires more efficiently. That is, she can implement the same slope λ_i at a lower cost. In effect, a higher $\bar{\lambda}$ reduces the cost of any given λ_i. This creates a strategic complementarity in addition to the more obvious one implied by the form of the game's payoff U_i.

Equilibrium Since the best response to linear-Gaussian profiles is linear-Gaussian (and continuous on the compact interval $[0,1]$), equilibria in linear-Gaussian strategies exist.[9] Moreover, for a large enough coordination motive γ and an appropriate range of information costs μ, the complementarities descibed above can create multiple linear-Gaussian equilibria (see Vives 2005). Proposition 1 characterizes these equilibria and describes their stability properties. The notion of stability used is defined below (see Manzano and Vives 2011).

[9] In fact (as is shown in Appendix B), such equilibria exist only if the prior is Gaussian.

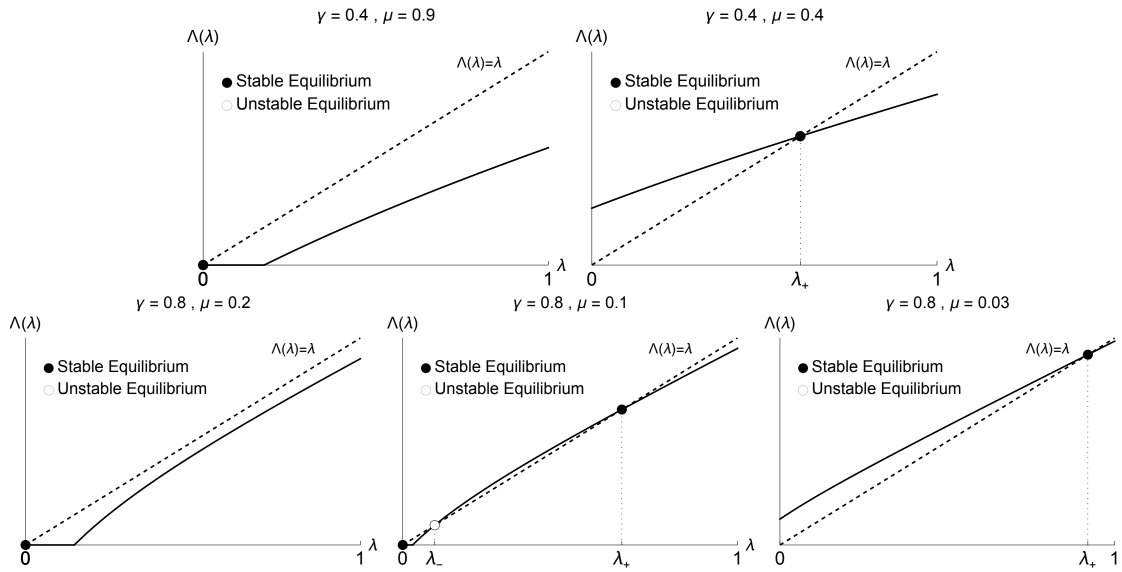


Figure 1: Examples of best responses and stability analysis for the slope λ of the average action in linear-Gaussian equilibrium ($\theta \sim N(0,1)$).

Definition 3. *A linear-Gaussian equilibrium is stable if its corresponding slope $\lambda^* \in [0,1]$ is a stable fixed point for the best-response function $\Lambda(\cdot)$, i.e., if $|\Lambda'(\lambda^*)| < 1$. It is unstable otherwise.*

Proposition 1. *Let p be Gaussian. Then, generically, the game (p, γ, μ) has either one or three linear-Gaussian equilibria. If it has three linear-Gaussian equilibria, then one of them has a slope $\lambda = 0$ and is stable. From the two equilibria with $\lambda > 0$, the one with the higher λ is stable, while the one with the lower λ is unstable.*

Proof. See Appendix A.1.

The equilibria of Proposition 1 (in particular, the stable equilibrium with positive λ) will serve as the benchmark in the sensitivity analysis of Section 5. Figure 1 plots the best-response function for selected (γ, μ) combinations. The mechanisms behind equilibrium multiplicity are briefly discussed in Section 6.

4 Equilibrium characterization with an arbitrary prior

Relaxing the normal prior assumption that was imposed in Section 3, Section 4 proceeds to address the general model laid out in Section 2. It begins by calculating best responses and proceeds with characterizing equilibria in continuous strategies.

4.1 Best responses

This paper studies players' behavior in equilibria where they follow strategies that have densities, termed *continuous strategies*. A formal definition follows.

Definition 4 (Continuous strategy). *A strategy of player i is* continuous *if the marginal $P_{a_i}(\cdot) = \int_{-\infty}^{+\infty} P_{a_i|\theta}(\cdot|\theta)p(\theta)\,\mathrm{d}\theta$ is absolutely continuous with respect to the Lebesgue measure.*

Note that $P_{a_i|\theta}(\cdot|\theta)$ being absolutely continuous for P_θ-almost all θ is sufficient for player i's strategy to be continuous, but it is not necessary. For example, the strategy that assigns probability mass 1 to the action $a_i = \theta$ is continuous: the marginal $P_{a_i} = P_\theta$ is absolutely continuous, even though $P_{a_i|\theta}(\cdot|\theta)$ is not absolutely continuous for any θ. In contrast, the constant-action strategy that assigns probability mass 1 to action $a_i = 0$ for all θ leads to P_{a_i} putting all probability mass to $a_i = 0$ and, therefore, is not continuous.

As will be promptly shown (Lemma 2), when information costs are low enough, continuous strategies are best responses to strategy profiles that satisfy certain smoothness conditions, termed *smooth, monotone, full-support profiles.*

Definition 5 (Smooth, monotone, full-support profile). *A strategy profile r is a* smooth, monotone, full-support profile *if*

1. *the profile's average action function $\bar{a}(\cdot)$ is analytic in its argument, and*

2. *$\bar{a}'(\theta) > -\frac{1-\gamma}{\gamma}$ for all $\theta \in \mathbb{R}$.*

The requirements of Definition 5 ensure that the best action function $b(\cdot)$ is strictly increasing, analytic, and bijective. Condition 2 is not too restrictive: it is satisfied for any increasing $\bar{a}(\cdot)$ but also allows for decreasing average action functions, as long as the decrease is not too fast (even though such behavior may not make intuitive sense).[10] Since in a smooth, monotone, full-support profile $b(\cdot)$ is bijective, it is also invertible. Let $\theta(\cdot) := b^{-1}(\cdot)$ denote the inverse of $b(\cdot)$ and g denote the PDF of the distribution that the best action follows. The PDF g is, then,

$$g(\cdot) = p\,(\theta(\cdot))\,\theta'(\cdot) \tag{7}$$

and is analytic. The variance of the best action (the variance of g) is denoted by σ_b^2.

The first result of the paper about general priors is Lemma 2. It provides necessary and sufficient conditions for the existence of a continuous best response to a smooth, monotone, full-support profile (see also Matějka and Sims 2010). In the result's statment—and throughout the paper—the Fourier and inverse Fourier transforms are denoted as

$$\begin{aligned}
\mathscr{F}_x[f(x)](\xi) &= \int_{-\infty}^{+\infty} f(x)\exp(-2\pi\imath x\xi)\,\mathrm{d}x \\
\text{and } \mathscr{F}_\xi^{-1}[F(\xi)](x) &= \int_{-\infty}^{+\infty} F(\xi)\exp(2\pi\imath x\xi)\,\mathrm{d}\xi
\end{aligned}$$

respectively ($\imath$ is the imaginary unit). Moreover, the shorthand notation with the hat operator $\hat{f}(\xi) := \mathscr{F}_x[f(x)](\xi)$ is occasionally used.

Lemma 2. *Consider a game (p, γ, μ). Let r_{-i} be a smooth, monotone, full-support strategy profile of player i's opponents. Player i has a continuous best response to r_{-i} if and only if*

$$R_i := \mathscr{F}_\xi^{-1}\left[\exp\!\left(\mu\pi^2\xi^2\right)\hat{g}(\xi)\right] \qquad \textit{is the PDF of a probability distribution.} \tag{8}$$

This continuous strategy is her unique best response and is given by

$$r_i(a_i|\theta) = \begin{cases} \delta(a_i - b(\theta)) & \textit{if } \mu = 0 \\ R_i(a_i)\frac{b'(\theta)}{p(\theta)}\frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{(a_i-b(\theta))^2}{\mu}\right) & \textit{if } \mu > 0 \end{cases}$$

where $\delta(\cdot)$ is Dirac's delta function and $R_i(a_i)$ is the marginal density of action a_i.

Proof. See Appendix D.1.

[10] In fact, in Proposition 5 it will be shown that such unintuitive behavior does not take place in equilibrium, even though the definition of smooth, monotone, full-support profile does not preclude it.

Lemma 2 shows that results known for individual decision making ("tracking") problems under rational inattention (Matějka and Sims 2010) can be applied to derive best responses in strategic environments, provided the variable being tracked is appropriately defined.[11]

It is worth noting that continuous strategies are best responses not only to profiles where (almost) all opponents follow continuous strategies. In particular, they can be best responses even when all opponents are using discrete strategies à la Jung et al. (2019) or Matějka and McKay (2015), where the probability assigned to each action is an analytic function of the fundamental. Moreover, the best response to a smooth, monotone, full-support profile is continuous as long as information costs are low enough. So, when information is not too expensive, continuous equilibria should be more likely. These two observations are formalized in the following proposition.

Proposition 3. *Let $\left\{P_{a_j|\theta}\right\}_{j\neq i}$ be a profile of player i's opponents' strategies where $P_{a_i|\theta}(A|\theta)$ is analytic in θ for all Lebesgue-measurable sets A and all $j \neq i$. Let also $\bar{a}(\theta)$ be well-defined and $\bar{a}'(\theta) > -\frac{1-\gamma}{\gamma}$ for all θ. Then, there exists a $\mu^* \geq 0$ such that player i's best response to $\bar{a}(\cdot)$ is continuous if $\mu \in [0, \mu^*)$ and discontinuous if $\mu > \mu^*$.*

Proof. See Appendix D.2.

4.1.1 Conditional normality

According to Lemma 2, player i's best response is continuous only if the best action $\boldsymbol{b}$, viewed as a random variable, can be written as the sum of two independent random variables, one of which follows a Gaussian with variance $\mu/2$.[12] That is, only if there exists a PDF R_i for which the sum of the random variables $\boldsymbol{a}_i \sim R_i$ and $\boldsymbol{\varepsilon}_i \sim N(0, \mu/2)$ follows the same distribution as $\boldsymbol{b}$ ($\boldsymbol{a}_i + \boldsymbol{\varepsilon}_i = \boldsymbol{b}$).[13] For this to be feasible, the "resolution" $\sqrt{\mu/2}$ of $\boldsymbol{\varepsilon}_i$ must be small enough.[14] In this best response, the marginal distribution of

[11]Note that $\hat{R}_i$ as calculated from eq. (8) might be the Fourier transform of a probability distribution with atoms. In that case, player i's best response is still given by Lemma 2, but it is no longer in continuous strategies.

[12]Throughout the paper boldface letters denote random variables.

[13] To see this, from eq. (8), one gets $\hat{g}(\xi) = \hat{R}_i(\xi)\ \exp\left(-\mu\pi^2\xi^2\right) = \hat{R}_i(\xi)\ \mathscr{F}_x\left[\frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{x^2}{\mu}\right)\right](\xi)$ and so, by the convolution theorem, g is the convolution of R_i and the Gaussian $N(0, \mu/2)$.

[14]A necessary condition is that R_i as calculated from (8) has a positive variance, i.e., that $\sigma_b^2 > \mu/2$. Another, stronger, necessary condition is that $\text{Var}(\boldsymbol{a}_i|b) > 0$ for all b.

the player's action is that of the "residual" variable $\boldsymbol{a}_i$. It follows that, when she best responds, player i's posterior belief on b is (i) normally distributed and (ii) independent of the prior distribution of the fundamental. Proposition 4 provides a formal statement.

Proposition 4. *In player i's continuous best response to a smooth, monotone, full-support strategy profile, her posterior belief about the best action b has a PDF given by*

$$\varrho_i(b|a_i) = \frac{1}{\sqrt{\pi\mu}} \exp\left(-\frac{(a_i - b)^2}{\mu}\right).$$

Proof. See Appendix D.3.

This is a result of two things: the quadratic-losses objective and the Shannon-entropy-based information costs. First, the quadratic-losses form of the objective function imposes ex-post optimality—i.e., that the action taken should be equal to the ex-post belief about the best action (see Observation 1)—and penalizes strategies according to the variance of the action a_i around b (see eq. (3)). So, when taking action a_i, player i's posterior belief about b should be centered at a_i and with the smallest variance possible. Now, among the family of distributions with full support on $\mathbb{R}$ which have a given mean a_i and variance σ^2, the Gaussian $N(a_i, \sigma^2)$ is the distribution with the maximum Shannon entropy (in this sense, the normal distribution is "informationally efficient"). Therefore, a normally distributed posterior is the "cheapest" one that achieves any given variance level.[15]

4.1.2 Expected action in a best response

Finally, Proposition 5 offers a reassuring result. It confirms that, in line with intuition, in a player's best response, her expected action follows the direction in which the best action and the fundamental move.

Proposition 5. *Let g be the distribution of the best action of a smooth, monotone, full-support strategy profile of player i's opponents that satisfies condition* (8)*. Then, in player i's best response,* $b \mapsto \mathbb{E}(\boldsymbol{a}_i|b)$ *and* $\theta \mapsto \mathbb{E}(\boldsymbol{a}_i|\theta)$ *are increasing functions.*

Proof. See Appendix D.6.

[15]The Online Appendix has a variant of this result for a broader class of games. Just as in Proposition 4, a player's posterior on b does not depend on the prior but only depends on the payoff function and the information cost. It is Gaussian iff payoffs are given by quadratic losses. See also Jung et al. (2019).

4.2 Equilibrium

Building on the results of Section 4.1, this section sets out to characterize the class of equilibria in which players use continuous strategies (see Definition 4).

Definition 6 (Smooth, monotone, full-support equilibrium)**.** *A strategy profile r is called a* smooth, monotone, full-support equilibrium *if (a) it is is a smooth, monotone, full-support profile (Definition 5), (b) it is an equilibrium (Definition 1), and (c) each player's strategy r_i is continuous (Definition 4).*

The following proposition characterizes the class of SMFE. It is the main technical contribution of the paper and enables the sensitivity analysis of Section 5.

Proposition 6. *Consider a game (p, γ, μ). The following two statements are equivalent*

(A) *$\theta(\cdot)$ is the inverse of the best action function and g is the PDF of the distribution of the best action in an SMFE.*

(B) *$\theta : \mathbb{R} \to \mathbb{R}$ is a strictly increasing bijection, $\mathscr{F}_{\xi}^{-1}[\exp(\mu\pi^2\xi^2)\hat{g}(\xi)]$ is a probability distribution,*

$$\theta(b) = b - \frac{\mu\gamma}{2(1-\gamma)} \frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b))), \quad \textit{and} \tag{9}$$

$$g(b) = p(\theta(b))\theta'(b). \tag{10}$$

Proof. See Appendix D.7.

While equation (9) holds independently of the prior and has to do with the way individuals acquire information, equation (10) forces the distribution g to be generated by the particular best action function (or, to be precise, its inverse) given the fundemental's prior distribution p, as seen in equation (7). According to Proposition 6, a distribution g generates a unique $\theta(\cdot)$ through (9). Similarly, a best action function with inverse $\theta(\cdot)$ generates a unique distribution g through (10). If these hold simultaneously, then an equilibrium has been identified.

In Appendix B it is shown that SMFE in which the best action is an affine function of the fundamental exist only if the prior is Gaussian. While Gaussian priors deliver tractable beauty contest models, the analyst should be careful when making this assumption, as her predictions may not be robust. Section 5 argues more to that point.

Although a general equilibrium existence theorem is not provided, Sections 3 and 5 demonstrate that SMFE do exist for certain classes of priors and for information costs μ bounded away from zero. With this in hand, the existence of SMFE for other families of priors and non-vanishing costs can be established through homotopy arguments.

Some properties of this class of equilibria are described in the Online Appendix. There, it is also shown that if players have heterogeneous information costs, the characterization result of Proposition 6 still holds, as long as all players have low enough information costs. One only needs to replace the cost μ in eq. (9) with the population-wide average cost $\bar{\mu}$. Finally, the Online Appendix characterizes the SMFE in a broader class of games. The absence of explicit payoff formulas in these more general games, though, limits the applicability of the characterization therein.

4.3 A method to address non-normal priors

Despite it being challenging to identify equilibria in the usual way (i.e. to find smooth, monotone, full support equilibria for a given parameter combination (p, γ, μ)), considerable progress can be made through following a "backwards" procedure. In particular, one can postulate some distribution g to be the PDF of the best action b in an SMFE and then, making use of equation (9), calculate (analytically or numerically) the prior distribution of the fundamental through[16]

$$p(\cdot) = g(b(\cdot))b'(\cdot).$$

Similarly, one can calculate the mapping $\alpha(\cdot)$ that gives the average action as a function of b (Lemma 9 in Appendix D.5) and the PDF of the equilibrium average action through

$$h(\cdot) = g\left(\alpha^{-1}(\cdot)\right)\left(\alpha^{-1}\right)'(\cdot).$$

A major challenge that arises during this process lies in confirming whether condition (8) holds, i.e., whether the calculated distribution of a player's action is, indeed, a

[16]To ensure that $b(\cdot)$ is strictly increasing (Definition 5), γ and μ should be low enough so as to make $\max_{b\in\mathbb{R}} (\log(g(b)))'' < 2(1-\gamma)/\mu\gamma$ (see eq. (9)).

(continuous) distribution. In order to overcome this problem, one can make use of best action distributions g which are conjugate for the normal distribution.

Since the equilibrium best action $\boldsymbol{b}$ is the sum of $\boldsymbol{a}_i \sim R_i$ and $\boldsymbol{\varepsilon}_i \sim N(0, \mu/2)$ (Section 4.1.1), when g is conjugate for the Gaussian, g and R_i belong to the same distribution family with R_i having the same mean as g and a variance reduced by $\mu/2$. In this way, confirming that (8) gives, indeed, the PDF of a probability distribution boils down to making sure that the parameters calculated for R_i fall within the allowed ranges for the particular distribution family.

4.3.1 Distribution families

Two families of distributions are used in the applications of Section 5: the *skew normal* ($SN(\chi)$)—where the parameter $\chi \in \mathbb{R}$ determines the skewness of the distribution—and equal-weight mixtures of two Gaussians of the same variance but whose centers are around β and $-\beta$ (termed *mixture normal* and denoted as $MN(\beta)$). These families are described in detail in Appendix C. Note that each of these distributions reduces to the standard normal, if the respective parameter is set to zero.

5 Equilibrium sensitivity to misspecifications of the prior

Employing the characterization result and the method of Section 4, this section studies how sensitive analysts' predictions are to misspecifications of the fundamental's distribution. Section 5.1 uses examples to show that moving away from the Gaussian prior assumption leads to new insights about how equilibrium actions and economic fundamentals are related. Sections 5.2 and 5.3 calculate error amplification factors for Gaussian misspecifications of Gaussians and non-Gaussians. Finally, Section 5.4 compares the two cases and shows that assuming a Gaussian prior leads to more severe errors when the true distribution of the fundamental is non-Gaussian than when it is an equally-misspecified Gaussian.

5.1 Qualitative implications of using misspecified priors

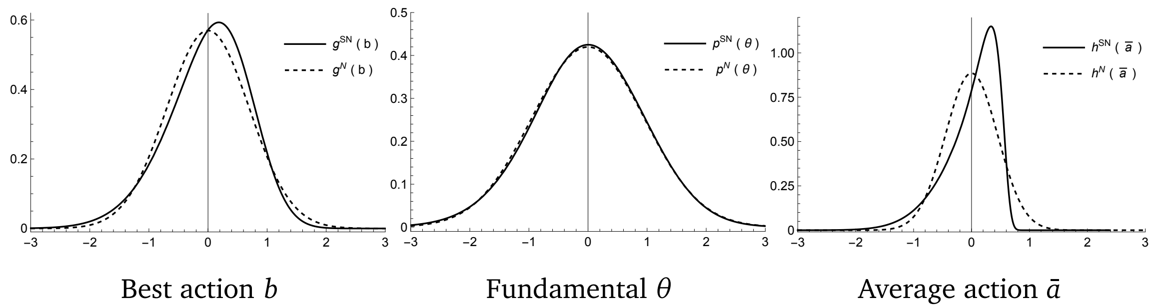
In applied theoretical research, using normal priors when real-world distributions of economic fundamentals are, in fact, non-normal will obviously lead to predictions that are not entirely accurate. One may wonder, though, to what extent a simplifying assumption of normality produces inaccurate results and, importantly, whether such a simplification leads to the researcher missing out on new insights, especially given the high tractability of models with normal distributions (Section 3). Using examples of distributions from Appendix C (summarized in Section 4.3.1), this section demonstrates that even when prior distributions are very close to the normal, distributions of equilibrium objects can be qualitatively different from those derived for the normal.

In order to illustrate these differences, consider two beauty contests both of which have a coordination motive $\gamma = 0.5$ and information cost $\mu = 0.35$. In the first beauty contest there is an SMFE in which the best action's distribution g^{SN} is the skew normal $SN(-2)$. Similarly, in the second beauty contest there is an SMFE in which the best action is distributed according to a normal distribution g^{N} such that $\mathbb{E}(g^{\mathrm{N}}) = \mathbb{E}(g^{\mathrm{SN}}) = 0$ and $\mathrm{Var}(g^{\mathrm{N}}) = \mathrm{Var}(g^{\mathrm{SN}})$. The two distributions can be seen in the left-hand-side panel of Figure 2a.

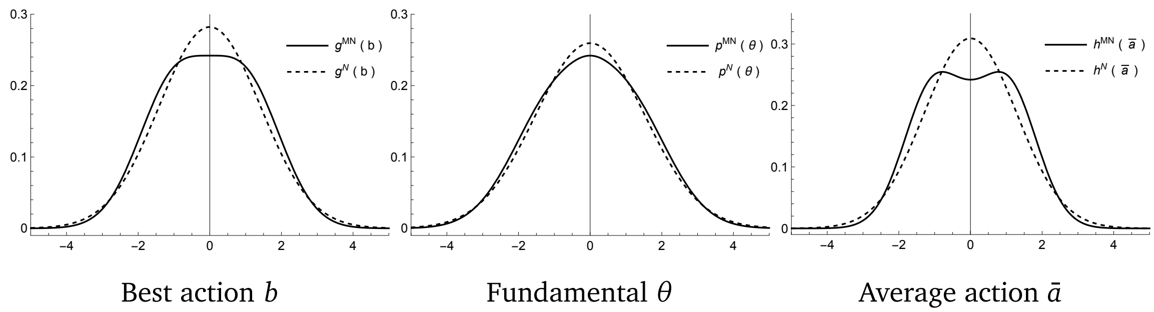
Using the method described in Section 4.3, one can calculate the prior distribution of the fundamental and the equilibirium distribution of the average action for each of the two beauty contests. These are seen in the middle- and right-hand-side panels of Figure 2a, respectively. From the diagrams it is clear that although the two priors are barely different, the equilibrium distributions of $\bar{a}$ differ significantly. In particular, the distribution of the best action in the SMFE with g^{SN} is highly skewed (the right tail almost disappears), even though the prior that gives rise to it is only slightly away from the Gaussian prior of the second beauty contest.

The preceding analysis shows that highly skewed market obervables do not necessarily reflect highly skewed fundamentals but can be the result of *small* fundamental non-normalities, amplified by strong coordination motives and costly information.

Take, for example, a situation based on the model of Hellwig and Veldkamp (2009): Competing firms choose their (log) price a_i trying to match a demand shock (the fundamental) in the presence of coordination motives (higher average price leads to higher best-response price from an individual firm). Now consider an analyst who wants to



(a) Skew normal g^{SN} and normal g^N with $\text{Var}(g^{SN}) = \text{Var}(g^N)$



(b) Mixture normal g^{MN} and normal g^N with $\text{Var}(g^{MN}) = \text{Var}(g^N)$

Figure 2: Comparison of distributions of prior and equilibrium objects between SMFE with non-normal and normal best-action distributions. In all cases $(\mu, \gamma) = (0.35, 0.5)$.

study the distribution of the demand shock θ in this market. As fundamentals are unobservable, what the analyst can observe are individual and (more realistically) aggregate data, such as the average price. After multiple observations, the analyst observes a highly skewed distribution of the average price and concludes that the underlying distribution of the fundamental has a similar property and is by no means close to a Gaussian. The example presented above, shows that this reasoning is heavily flawed and that, in fact, the fundamental can be very close to a Gaussian.

Naïvely attempting to draw conclusions about the (unobservable) distribution of the fundamental from observed market behavior may lead to over-estimating its skewness and under-estimating its variance. Consequently, this can exacerbate the issues that statistical inference with skewed distributions brings about (e.g. constructing confidence intervals). These effects become stronger as information costs or coordination motives increase.

The results of the same exercise with a mixture normal distribution of the best ac-

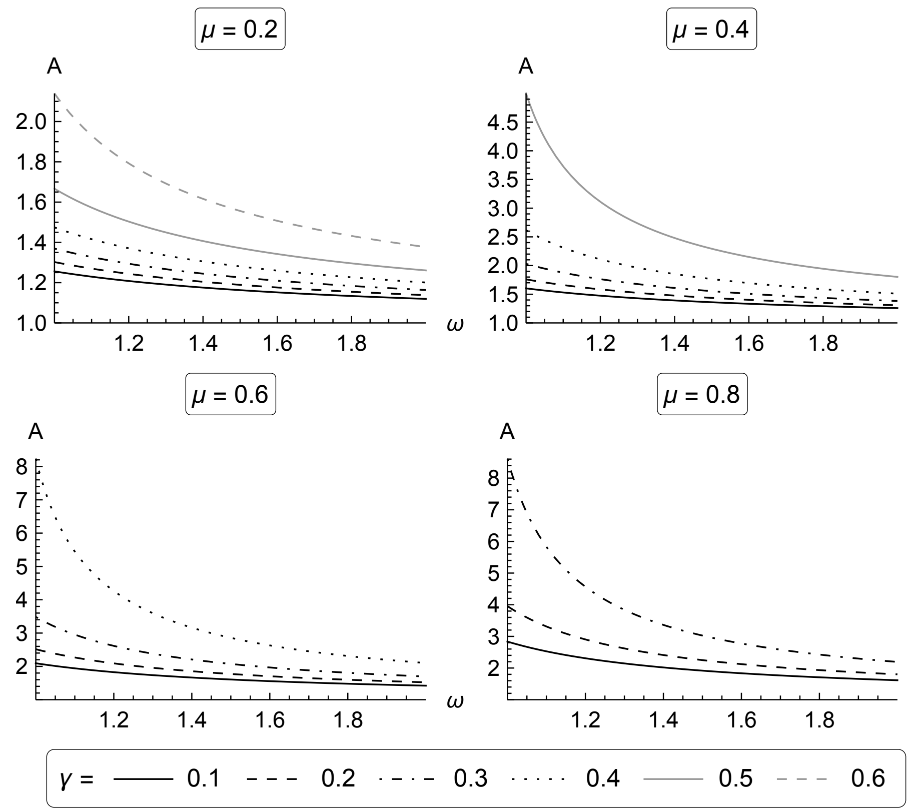


Figure 3: Error amplification for Gaussian misspecifications. The "true" prior has variance $\sigma^2 = \omega^2 > 1$, while the misspecified prior has variance $\tilde{\sigma}^2 = 1$. Curves are drawn only for (μ, γ) combinations for which both games have a linear-Gaussian equilibrium with $\lambda > 0$ (i.e., $\mu < 2(1-\gamma)^2$).

tion ($b \sim MN(1)$) are shown in Figure 2b. A similar observation can be made in this case: One should not conclude that the distribution of the fundamental is bimodal just because the distribution of the equilibrium average action is bimodal. The remainder of Section 5 aims to quantify the errors that prior misspecifications create and to analyze what true distributions of the fundamental lead to larger error amplifications.

5.2 Error amplification in a linear-Gaussian world

In the case of Gaussian priors and linear-Gaussian equilibria, the error amplification factor

$$A = \frac{\left\|h - \tilde{h}\right\|_1}{\|p - \tilde{p}\|_1}$$

of Section 2.1 can be calculated in closed form (see Appendix A.3). The analyst bases her predictions on the (mistaken) assumption that the distribution of the fundamental $\tilde{p}$ is the Gaussian $N(0, \tilde{\sigma}^2)$, while the true distribution p is $N(0, \sigma^2)$. The predictions $h, \tilde{h}$ associated the two priors are assumed to be the respective stable linear-Gaussian equilibria with positive slope λ_+ (see Proposition 1). Since the error amplification factor depends only on the quotient of the two standard deviations, the single parameter $\omega := \sigma/\tilde{\sigma}$ is used in what follows.

When information is costless ($\mu = 0$), in the unique equilibrium, the (linear) average action function has slope $\lambda = 1$. Thus, the distribution h of the equilibrium average action is equal to the distribution p of the prior for both the correct and the misspecified model. This means that any misspecifications of p translate to misspecifications of h at a one-to-one rate and, therefore, $A = 1$.

Figure 3 plots A values for varying ω in different (γ, μ) regimes, assuming $\tilde{\sigma} = 1$.[17] The figure shows that as ω increases, mistakes in the prior are amplified less (A decreases). However, this does not mean that the mistakes of the predictions get smaller in absolute value but, rather, that mistakes in predictions are increasing at a slower rate than mistakes in the prior. Importantly, in the presence of information costs, mistakes in the prior are always amplified ($A > 1$), moderately for small values of μ and γ but quickly increasing as these parameters increase.[18]

The intuition is the following. As each player faces higher fundamental uncertainty (i.e., σ increases), she suffers more severe losses from not (or from under-) acquiring information. That is, her marginal benefit from acquiring information increases. Now,

[17] Figure 3 focuses on values of $\omega > 1$. The same analysis can also be conducted for $\omega < 1$, as long as ω remains large enough for the game $\left(N\left(0, (\omega\tilde{\sigma})^2\right), \gamma, \mu\right)$ to have an equilibrium with positive slope λ (see Proposition 1).

[18] Note that as $\omega \to 1$, both the numerator and the denominator in the definition of A tend to zero. So, even though they tend to the same value, their quotient does not necessarily tend to 1. The amplification factor is equal to 1 only in the absense of information costs ($\mu = 0$).

as player i's information costs are convex in λ_i (see Appendix A.1), a higher marginal benefit from information implies that—in her best response—the player acquires more information as σ increases and her action becomes more responsive to the fundamental (λ_i increases). Since the game is one of strategic complements, the comparative statics of the "largest" equilibrium's slope λ_+ follow the same direction (see, e.g., Milgrom and Roberts 1994). So, the standard deviation of the equilibrium marginal-action distribution $\sigma_h = \lambda_+ \sigma$ reacts more-than-proportionally to changes of the prior standard deviation σ. This leads to any misspecification of p being amplified in h, i.e., to $A > 1$.

Overall, this shows that the predictions of such models are sensitive to prior distribution misspecifications, even when the true prior is Gaussian.

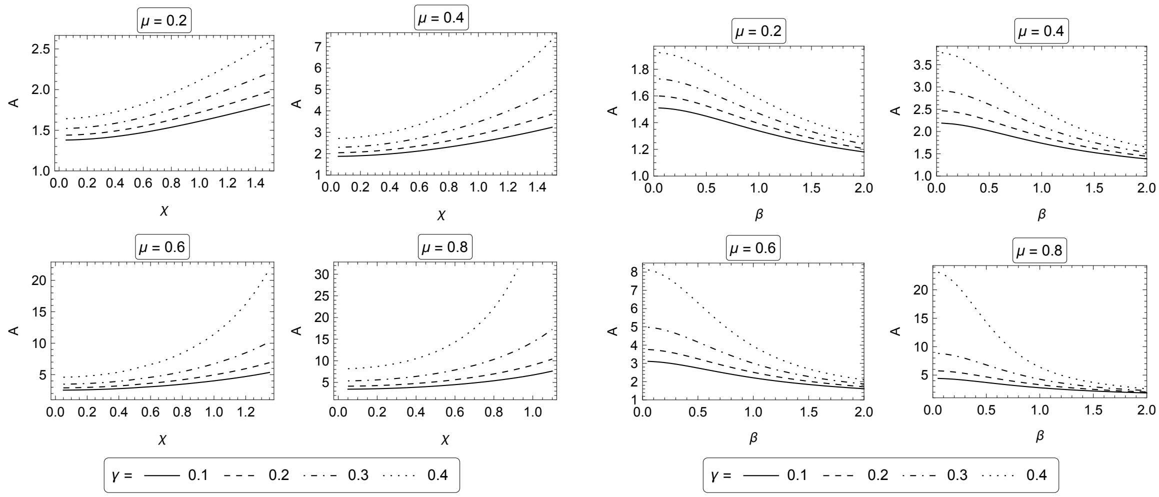
5.3 Error amplification with non-Gaussians

With the method of Section 4.3 and by using numerical techniques, one can conduct sensitivity analyses like the one in Section 5.2, even when closed-form solutions cannot be obtained. For this analysis, a non-Gaussian distribution is postulated to be the true best-action distribution g in an SMFE, under coordination and cost parameters (γ, μ) (with fundamental and average-action distributions p and h, respectively). Then, in the spirit of Section 2.1, an analyst who knows the values of γ and μ is assumed to make her predictions according to the misspecified (linear-Gaussian) SMFE whose best-action distribution $\tilde{g}$ is the Gaussian with $\text{Var}(\tilde{g}) = \text{Var}(g)$ (with fundamental and average-action distributions $\tilde{p}$ and $\tilde{h}$, respectively).

As in Section 5.2, the error amplification factor A is used to quantify how sensitive predictions are to the analyst's mistake of using a Gaussian prior instead of the correct, non-Gaussian one. Figure 4 presents the results of numerical calculations of A for the two families of g distributions described in Appendix C.[19]

The figure shows that higher error amplification factors are reached as information costs and the coordination motive increase. In the case of skew normal best-action distributions, larger deviations from the Gaussian increase the error amplification factor of A, which can reach values exceeding 20 for highly-skewed g. In contrast, in the case of mixture normal distributions g, the error amplification factor begins at relatively high

[19] Similar to the Gaussian misspecification case, when $\chi \to 0$ or when $\beta \to 0$, both the numerator and the denominator in the definition of A tend to zero. So, their quotient does not necessarily tend to 1.



(a) Skew normal g (b) Mixture normal g

Figure 4: Error amplification (A) for Gaussian misspecifications ($\tilde{p}$) of non-Gaussian distributions (p). In all comparisons, $\mathrm{Var}(\tilde{g}) = \mathrm{Var}(g)$.

values but decreases as g deviates further from the Gaussian (while always exceeding unity). This is due to the $\|\tilde{p} - p\|_1$ distance increasing faster than $\left\|\tilde{h} - h\right\|_1$ (see also Section 5.2 and fig. 3). Echoing the qualitative results of Section 5.1, Figure 4 quantitatively demonstrates that errors made by incorrectly assuming a Gaussian prior in place of a non-Gaussian one can lead to much larger errors in an analyst's predictions, especially if fundamentals are skewed.

5.4 Comparison between Gaussian misspecifications of Gaussians and non-Gaussians

Say that an analyst makes predictions under the mistaken assumption that the fundamental is distributed according to the Gaussian distribution $\tilde{p}$. Under what true distributions p does her mistake lead to larger errors in her predictions? Note that direct comparisons between Figures 3 and 4 cannot give a satisfactory answer, as the comparisons in the two figures are between different pairs of distributions p and $\tilde{p}$.

To make the comparison meaningful, the analyst is assumed to be using *the same* Gaussian $\tilde{p}$ for her predictions, while the two scenarios compared are: (a) the true distribution of the fundamental is a non-Gaussian p and (b) the true distribution of the fundamental is a Gaussian $\check{p}$. Importantly, the distributions p and $\check{p}$ are such that the analyst's assumption $\tilde{p}$ is equally wrong (according to the $\|\cdot\|_1$ measure) in the two scenarios, i.e.,

$$\|\check{p}-\tilde{p}\|_1 = \|p-\tilde{p}\|_1. \tag{11}$$

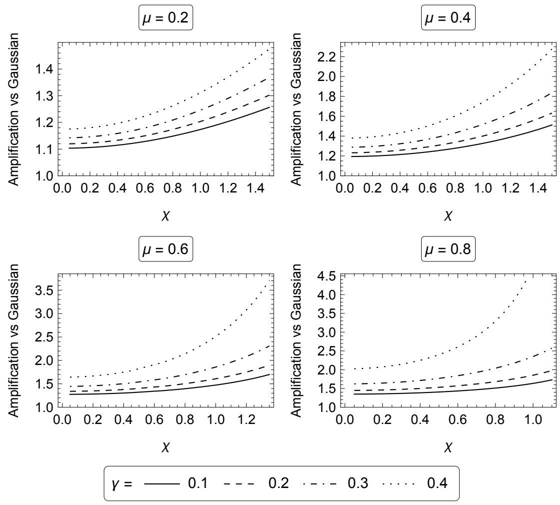
To operationalize this approach computationally, starting from a postulated non-Gaussian distribution g of the best action in an SMFE, the related p and h distributions are calculated, along with the misspecified distributions $\tilde{g}$, $\tilde{p}$ and $\tilde{h}$, exactly as in Section 5.3. Then, using the distance $\|p-\tilde{p}\|_1$, the distribution $\check{p}$ is defined to be the Gaussian that is as "far" from $\tilde{p}$ as p is (eq. (11)).[20] Finally, $\check{h}$ is calculated as the distribution of the average action in the linear-Gaussian equilibrium of $\check{p}$ with the steepest slope λ (see Proposition 1). The extent to which the error of the Gaussian misspecification of a non-Gaussian is larger than that of a Gaussian is quantified by comparing the error amplification factors in the two scenarios:

$$\Omega = \frac{A}{\check{A}} = \frac{\left\|h-\tilde{h}\right\|_1 / \|p-\tilde{p}\|_1}{\left\|\check{h}-\tilde{h}\right\|_1 / \|\check{p}-\tilde{p}\|_1} = \frac{\left\|h-\tilde{h}\right\|_1}{\left\|\check{h}-\tilde{h}\right\|_1}.$$

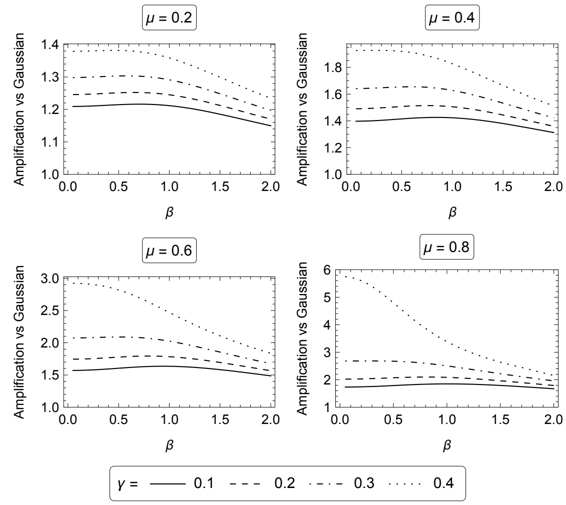
Larger values of Ω indicate that a quantitatively equal error in the measurement of the distribution of the fundamental leads to larger errors in the prediction of the equilibrium average action distribution when the true distribution is non-Gaussian than when the true distribution is Gaussian. The results of the calculations for the families of g distributions of Appendix C are seen in Figure 5.

The figure shows that the amplification rates follow the same patterns as those of Figure 4, even compared to amplification rates of equally-sized misspecifications of Gaussians. In particular, non-Gaussian fundamental distributions are always prone to larger errors than Gaussians. From an applied point of view, this means that analysts should be cautious when they have reasons to believe that fundamentals might not be Gaussian, as their predictions from using the Gaussian model can bear errors much larger than if they were to just use the wrong variance of a Gaussian.

[20] As the L^1 distance between two Gaussians with the same mean depends only on the ratio between their standard deviations (see Appendix A.2), for any given Gaussian $\tilde{p}$ (with standard deviation $\tilde{\sigma}$) and distance $\delta \in (0,2)$ there are two Gaussians $\check{p}$ that solve $\|\tilde{p}-\check{p}\|_1 = \delta$. These solutions have standard deviations $\omega\tilde{\sigma}$ and $\tilde{\sigma}/\omega$ for some $\omega > 0$. To guarantee the existence of a linear-Gaussian equilibrium for $\check{p}$, the solution of eq. (11) with the larger standard deviation is used to define $\check{p}$.



(a) Skew normal g



(b) Mixture normal g

Figure 5: Error amplification comparisons ($\|h-\tilde{h}\|_1/\|\check{h}-\tilde{h}\|_1$) between Gaussian misspecifications ($\tilde{p}$) of non-Gaussian (p) and Gaussian ($\check{p}$) distributions. In all comparisons, $\mathrm{Var}(\tilde{g}) = \mathrm{Var}(g)$ and $\|\check{p}-\tilde{p}\|_1 = \|p-\tilde{p}\|_1$.

6 Discussion

This paper studied beauty contests played by rationally inattentive agents. It demonstrated that while assuming a Gaussian prior makes such models tractable, their predictions are very sensitive to misspecifications of the prior. What follows discusses and relates the paper's results to the ones found in other studies.

Equilibrium sensitivity Jung et al. (2019) make the point that small differences in the prior of decision-theoretic problems of rational inattention can lead to very different behavior of the agent. In particular, the agent may switch from continuous to discontinuous strategies in problems whose priors are very close. In Section 5 it was shown that in *games* with rationally inattentive agents problems with very similar priors can lead to very different equilibrium behavior even if the equilibria in both problems are in continuous strategies. The reason for the high sensitivity of solutions to the prior in Jung et al. (2019) is the information cost being too high for continuous strategies to be optimal in one of the problems (see Proposition 3 in this paper), or the distribution having fat tails. In contrast, the driving force behind the similar result of Section 5 lies in the coordination motive working together with the information cost. Even small asymmetries in the prior can lead to players having a preference for one side of the distribution over the other when they coordinate (since they cannot follow the fundamental closely because of information costs). Similarly, the distribution of the equilibrium average action can exhibit different features from that of the fundamental (e.g., bimodality).

Equilibrium multiplicity Entropy-related information costs, as those used in this paper, can lead to multiple equilibria (e.g. Hellwig, Kohls, and Veldkamp 2012). In the present model, multiple linear-Gaussian equilibria arise for *intermediate* values of μ (bounded away from zero for any fixed $\gamma > 1/2$). To understand why, recall that while higher information costs disincentivize information acquisition, they also strenghthen strategic complementarities (Section 3). Equilibrium multiplicity, then, requires information to be costly enough for the strategic complementarity to be strong, but not too costly, so that an equilibrium with positive information acquisition can be sustained (see also Hébert and La'O (2020) and Myatt and Wallace (2012) for similar analyses). The largest and the smallest equilibria are stable (Echenique 2002; Vives 1990), while there is an intermediate unstable equilibrium.[21]

With flexible information acquisition technology, when information is cheap, players obtain more of it and the game gets closer to a complete-information one. Yang (2015) uses such a technology to study a two-player coordination game. The complete-information game has multiple equilibria for a range of realizations of the random variable and this multiplicity is recovered when information costs are low. Both in this paper and in Yang (2015), as information costs vanish, the equilibrium structure of the complete-information game is recovered: a unique equilibrium in the present case, multiple in Yang's.

According to Morris and Yang (forthcoming), what drives this result is the entropy-reduction cost function. In their setting—which has multiple equilibria under complete information—continuous stochastic choice breaks when it is sufficiently easy to distin-

[21] See also Manzano and Vives (2011) who study equilibrium stability in a general CARA-normal model of financial markets. In their model strategic decisions are complements/substitutes at different points of the strategy space, whereas in this paper information acquisition decisions are always strategic complements.

guish nearby states, and multiple equilibria appear.[22] Under entropy costs, nearby and far away states are equally hard to distinguish, making discontinuous stochastic choice feasible. In turn, this leads to the equilibrium structure of the complete-information game being recovered as information costs vanish. In contrast, in the global-games context (Carlsson and van Damme 1993), arbitrarily accurate signals lead to unique equilibrium selection, as stochastic choice is continuous (see Vives 2005, sec. 7.2 for an analysis of the driving forces behind this result).

Linear equilibria The notion of a linear equilibrium is often encountered in existing literature studying linear-quadratic games (see for example Angeletos and Pavan 2007; Morris and Shin 2002; Myatt and Wallace 2012). In linear equilibria each player takes an action that is a linear combination of the messages she receives from (potentially) different sources. When signal noises and the prior follow normal distributions—which is the common modelling choice in the aforementioned literature—then any linear equilibrium is a linear-Gaussian one in the sense of Section 3. As shown in Appendix B, linear tractability of equilibria when players are rationally inattentive is heavily dependent on this very assumption. Importantly, Proposition 6 and the method of Section 4.3 can help identify or approximate equilibria even when the prior is not normal.

Conditionally correlated signals Hellwig and Veldkamp (2009) point out that strategic complementarities lead to complementarities in players' information acquisition decisions. Moreover, one can think of situations where a player may want to have information about other players' signal realizations or may even want other players to have information about her own realization (as in Kozlovskaya 2018, for example). The model presented here does not allow for such correlation, as the only information players can obtain is about the fundamental and not about others' signal realizations (signals are always conditionally independent). However—as Denti (2020) argues—when all players are "small" and aggregate behavior is all that matters, incentives to learn about others' signal realizations disappear, since aggregate behavior becomes a deterministic function of the fundamental.

[22] Goryunov and Rigos (2020) test the predictions of this model in a laboratory experiment.

Efficiency Individual players do not consider how their own information acquisition affects the average action and may, therefore, under-acquire information, compared to the social optimum.[23] This externality can give rise to inefficient equilibria as, for example, the unstable linear-Gaussian equilibrium with the smaller positive slope λ_{-} (see Proposition 1 and fig. 1). However, Hébert and La'O (2020) argue that (a discretized version of) any game (p, γ, μ) is guaranteed to have efficient equilibria, given the mutual-information-based cost function and the linear-quadratic payoff function used in this paper. Following the intuition behind the externality, efficient equlibria are the ones in which players acquire the most information. Moreover, continuous strategies convey more information than "similar" strategies that assign probability mass to only a discrete subset of actions (see Matějka and Sims 2010). Consequently, if a game admits equilibria in continuous strategies (SMFE), efficent equilibria should be in this class.

Appendix

A Calculations for linear-Gaussian equilibria

A.1 Proof of Proposition 1

Consider a profile in which all of player i's opponents use linear-Gaussian strategies. Then, the average-action function to which player i best responds is defined through $\bar{a}(\theta) = \int_0^1 \lambda_j \theta \,\mathrm{d} = \bar{\lambda}\theta$ and the best-action function through $b(\theta) = \kappa\theta$ with $\kappa := (1-\gamma) + \gamma\bar{\lambda}$. So, player i's optimization problem boils down to tracking the variable $b \sim N(0, (\kappa\sigma)^2)$ and her optimal strategy is linear-Gaussian (see Sims 2003; Jung et al. 2019).

Now, when using a linear-Gaussian strategy, player i's posterior beliefs on the fun-

[23] Regarding this externality, see also the discussion on strategic complementarity through information costs in Section 3.

damental and on the best action follow Gaussian distributions. From Bayes's law:

$$\theta|a_i \sim N\left(\frac{\sigma^2\lambda_i a_i}{\lambda_i^2\sigma^2+\sigma_{a_i}^2}, \frac{\sigma^2\sigma_{a_i}^2}{\lambda_i^2\sigma^2+\sigma_{a_i}^2}\right) \qquad b|a_i \sim N\left(\frac{\kappa\sigma^2\lambda_i a_i}{\lambda_i^2\sigma^2+\sigma_{a_i}^2}, \frac{\kappa^2\sigma^2\sigma_{a_i}^2}{\lambda_i^2\sigma^2+\sigma_{a_i}^2}\right)$$

Payoffs

The expected value of the objective of a player i who uses a linear-Gaussian strategy (λ_i, σ_i^2) against a best-action slope κ can be calculated to be

$$U_i = \mathbb{E}[-(a_i - b)^2; (\lambda_i, \sigma_{a_i}^2)] = -\sigma_{a_i}^2 - (\kappa - \lambda_i)^2\sigma^2. \tag{12}$$

Since both the prior and all posterior beliefs are Gaussian, the cost of a linear-Gaussian strategy $(\lambda_i, \sigma_{a_i}^2)$ is (see eq. (4))

$$C_i = \frac{\mu}{2}\log\left(\frac{\sigma_{\text{prior}}^2}{\sigma_{i,\text{posterior}}^2}\right) = \frac{\mu}{2}\log\left(1 + \frac{\lambda_i^2\sigma^2}{\sigma_{a_i}^2}\right). \tag{13}$$

Ex-post consistency (Observation 1) imposes

$$\sigma_{a_i}^2 = \lambda_i(\kappa - \lambda_i)\sigma^2 \tag{14}$$

on player i's strategy. Substituting eq. (14) into eq. (12) and eq. (13) makes player i's objective:

$$U_i - C_i = -\kappa(\kappa - \lambda_i)\sigma^2 - \frac{\mu}{2}\log\left(\frac{\kappa}{\kappa - \lambda_i}\right)$$

and reduces the decision problem of player i to simply choosing $\lambda_i \in [0, 1]$. The solution to the maximization problem leads to the best-response function (after substituting $\kappa = (1-\gamma) + \gamma\bar{\lambda}$):

$$\Lambda(\bar{\lambda}) := \begin{cases} ((1-\gamma)+\gamma\bar{\lambda}) - \frac{\mu}{2((1-\gamma)+\gamma\bar{\lambda})\sigma^2} & \text{if } \mu < 2((1-\gamma)+\gamma\bar{\lambda})^2\sigma^2 \\ 0 & \text{otherwise} \end{cases}.$$

The equilibrium condition that $\Lambda(\bar{\lambda}) = \bar{\lambda}$, leads to the equilibrium slopes being

$$\lambda_+ = \frac{1}{2\gamma}\left(2\gamma - 1 + \sqrt{1 - \frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right) \qquad \lambda_- = \frac{1}{2\gamma}\left(2\gamma - 1 - \sqrt{1 - \frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right). \tag{15}$$

The number of equilibria varies, depending on the model parameters:

1. If either (a) $\gamma \leq \frac{1}{2}$ and $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$ or (b) $\gamma > \frac{1}{2}$ and $\sigma^2 \geq \frac{\mu}{2(1-\gamma)^2}$, then there is a unique equilibrium with $\lambda^* = \lambda_+$ ($\lambda^* = 1 - \frac{\mu}{2\sigma^2}$ when $\gamma = 0$).

2. If $\gamma > \frac{1}{2}$ and $\sigma^2 = \frac{2\mu\gamma}{(1-\gamma)}$, then $\lambda_+ = \lambda_-$ and there are two equilibria with slopes $\lambda^* \in \{\lambda_+, 0\}$ (this case is not generic, though).

3. If $\gamma > \frac{1}{2}$ and $\sigma^2 \in \left(\frac{2\mu\gamma}{1-\gamma}, \frac{\mu}{2(1-\gamma)^2}\right)$ then there are three equilibria with slopes $\lambda^* \in \{\lambda_+, \lambda_-, 0\}$.

4. If either (a) $\gamma \leq \frac{1}{2}$ and $\sigma^2 \leq \frac{\mu}{2(1-\gamma)^2}$ or (b) $\gamma > \frac{1}{2}$ and $\sigma^2 < \frac{2\mu\gamma}{(1-\gamma)}$, then there is a unique equilibrium with slope $\lambda^* = 0$.

Stability

The stability condition is $|\Lambda'(\lambda^*)| < 1$. By evaluating $\Lambda'(\lambda)$ at the different equilibrium slopes λ, the stability results are obtained for the relevant cases (excluding the non-generic case 2.):

1. The unique equilibrium is stable.

3. The equilibria with slopes $\lambda = 0$ and $\lambda = \lambda_+$ are stable. The equilibrium with slope $\lambda = \lambda_-$ is unstable.

4. The unique equilibrium is stable. □

A.2 Calculation of distance between two Gaussians

Let f_1 and f_2 be the PDFs and F_1 and F_2 be the CDFs of two Gaussians with mean 0 and variances σ_1^2 and σ_2^2, respectively ($\sigma_2 > \sigma_1$, without loss). There is a unique point x^* such that $x < x^*$ implies that $f_2(x) > f_1(x)$ (see Figure 6). This point is given by

$$x^* = -\sqrt{\frac{2\log\left(\frac{\sigma_2}{\sigma_1}\right)\sigma_1^2\sigma_2^2}{\sigma_2^2 - \sigma_1^2}}.$$

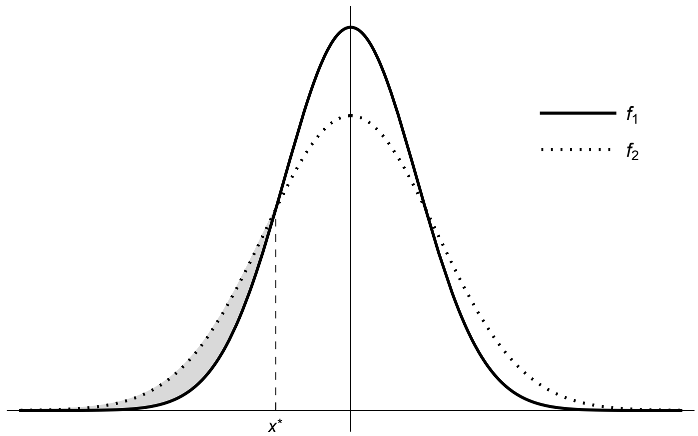


Figure 6: The L^1 distance between Gaussians. The shaded area is equal to $\|f_1 - f_2\|_1/4$.

Due to the symmetry of f_1 and f_2, $F_1(0) = F_2(0) = 1/2$. So, the L^1 distance between f_1 and f_2 is:

$$\begin{aligned}\|f_1 - f_2\|_1 &= 2\int_{-\infty}^{0} |f_1(x) - f_2(x)|\,\mathrm{d}x = 2\left(\int_{-\infty}^{x^*} (f_2(x) - f_1(x))\,\mathrm{d}x + \int_{x^*}^{0} (f_1(x) - f_2(x))\,\mathrm{d}x\right)\\ &= 4\,(F_2(x^*) - F_1(x^*)) = 2\left(\operatorname{erf}\left(\frac{x^*}{\sigma_2\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x^*}{\sigma_1\sqrt{2}}\right)\right)\end{aligned}$$

Substituting x^* and using the fact that erf is an odd function, calculations yield

$$\begin{aligned}\|f_1 - f_2\|_1 &= 2\left(\operatorname{erf}\left(-\sqrt{\frac{\log\left(\frac{\sigma_2}{\sigma_1}\right)\sigma_1^2}{\sigma_2^2 - \sigma_1^2}}\right) - \operatorname{erf}\left(-\sqrt{\frac{\log\left(\frac{\sigma_2}{\sigma_1}\right)\sigma_2^2}{\sigma_2^2 - \sigma_1^2}}\right)\right)\\ &= 2\left(\operatorname{erf}\left(\sigma_2\sqrt{\frac{\log\left(\frac{\sigma_2}{\sigma_1}\right)}{\sigma_2^2 - \sigma_1^2}}\right) - \operatorname{erf}\left(\sigma_1\sqrt{\frac{\log\left(\frac{\sigma_2}{\sigma_1}\right)}{\sigma_2^2 - \sigma_1^2}}\right)\right).\end{aligned}$$

Now, if $\sigma_2 = \omega\sigma_1$ with $\omega > 1$, the formula for $L(\omega) := \|f_1 - f_2\|_1$ is obtained.

$$L(\omega) = 2\left(\operatorname{erf}\left(\omega\sqrt{\frac{\log\omega}{\omega^2 - 1}}\right) - \operatorname{erf}\left(\sqrt{\frac{\log\omega}{\omega^2 - 1}}\right)\right).$$

As the distance depends only on the ratio between σ_1 and σ_2, it must be that $L(\omega) = L(1/\omega)$. So, a more general formula, which holds for any $\omega > 0$, is

$$L(\omega) = 2\left|\operatorname{erf}\left(\omega\sqrt{\frac{\log\omega}{\omega^2 - 1}}\right) - \operatorname{erf}\left(\sqrt{\frac{\log\omega}{\omega^2 - 1}}\right)\right|. \tag{16}$$

A.3 Calculation of error amplification for Gaussians

The misspecified prior $\tilde{p}$ is assumed to have a standard deviation of $\tilde{\sigma}$, while the true prior p is assumed to have a standard deviation of $\omega\tilde{\sigma}$. The ratio between the standard deviations of the two resulting equilibrium average-action distributions is

$$\omega_h = \frac{\sigma_h}{\tilde{\sigma}_h} = \frac{\sigma\lambda_+}{\tilde{\sigma}\tilde{\lambda}_+} = \frac{\omega\tilde{\sigma}\lambda_+}{\tilde{\sigma}\tilde{\lambda}_+} = \omega\frac{\sqrt{1-\frac{2\mu\gamma}{(1-\gamma)\omega^2\tilde{\sigma}^2}}+2\gamma-1}{\sqrt{1-\frac{2\mu\gamma}{(1-\gamma)\tilde{\sigma}^2}}+2\gamma-1},$$

and the error amplification is

$$A = \frac{L(\omega_h)}{L(\omega)}$$

with $L(\cdot)$ as defined in eq. (16).

B Aggregately affine equilibria

The conditions of Proposition 6 imply that in an SMFE, the best action function satisfies

$$b(\theta) = \theta + \frac{\mu\gamma}{2(1-\gamma)}\frac{1}{b'(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\log\left(\frac{p(\theta)}{b'(\theta)}\right)\right) \tag{17}$$

and the average action function satisfies

$$\bar{a}(\theta) = \theta + \frac{\mu}{2(1-\gamma)(1+\gamma(\bar{a}'(\theta)-1))}\left(\frac{p'(\theta)}{p(\theta)} - \frac{\gamma\bar{a}''(\theta)}{1+\gamma(\bar{a}'(\theta)-1)}\right), \tag{18}$$

neither of which is possible to solve in the general case. Instead, one can use the above equations to characterize the SMFE of a specific class: those in which the equilibrium average action (and, thus, also the equilibrium best action) is affine in θ, like the linear-Gaussian equilibria of Section 3. These are formally defined as follows.

Definition 7 (Aggregately affine equilibrium)**.** *An SMFE will be called an* aggregately affine equilibrium (AAE) *if the best action function has the form* $b(\theta) = \kappa\theta + d$ *for some constants* $\kappa > 0$ *and* $d \in \mathbb{R}$.

The following proposition gives a necessary and sufficient condition for AAE to exist.

Proposition 7. *Consider a game* (p, γ, μ) *with* $\mu > 0$ *and* $\gamma > 0$*. The following statements are equivalent:*

(A) (p,γ,μ) *admits an aggregately affine equilibrium.*

(B) p *is the PDF of a normal distribution, and*

(i) either $\gamma \le \frac{1}{2}$ *and* $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$

(ii) or $\gamma > \frac{1}{2}$ *and* $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$.

Proof. "(*A*) ⇒ (*B*)"

In an AAE the best action function is given by $b(\theta) = \kappa\theta + d$. So, $b'(\theta) = \kappa$ and $b''(\theta) = 0$ for all θ. Moreover, an AAE is an SMFE, so $b(\cdot)$ should satisfy (17). From equation (17) one obtains:

$$\kappa\theta + d = \theta + \frac{\mu\gamma}{2(1-\gamma)}\frac{1}{\kappa}\frac{\mathrm{d}}{\mathrm{d}\theta}\log p(\theta).$$

And thus,

$$\log p(\theta) = \int \frac{2(1-\gamma)\kappa}{\mu\gamma}((\kappa-1)\theta + d)\,\mathrm{d}\theta + C$$

where $C \in \mathbb{R}$ is an integrating constant. It will have to be chosen so that the condition $\int_{-\infty}^{+\infty} p(\theta)\,\mathrm{d}\theta = 1$ is satisfied. From the previous equation:

$$\log p(\theta) = \frac{(1-\gamma)\kappa}{\mu\gamma}((\kappa-1)\theta^2 + 2d\theta) + C.$$

Completing the square in the brackets and taking the exponential of both sides one obtains:

$$p(\theta) = \exp(C')\exp\left(\frac{(1-\gamma)\kappa(\kappa-1)}{\mu\gamma}\left(\theta - \frac{d}{1-\kappa}\right)^2\right)$$

for some other constant C'. Now, for $\int_{-\infty}^{+\infty} p(\theta)\,\mathrm{d}\theta = 1$ to be satisfied, it has to be that $\kappa \in (0,1)$, otherwise the resulting p will not be integrable. It is clear that—for an appropriate selection of C'—the previous expression is a normal distribution with a mean $\theta_0 = d/(1-\kappa)$ and variance

$$\sigma^2 = \frac{\mu\gamma}{2(1-\gamma)\kappa(1-\kappa)}.$$

More than that, since in an AAE it has to be that $\sigma_b^2 > \mu/2$, one gets that $\kappa^2\sigma^2 > \mu/2$ i.e. that $\sigma^2 > \mu/2\kappa^2$. Using this together with the above equation, one gets that $\kappa > 1-\gamma$. So, a lower bound for the value of σ^2 is given by the solution to the problem

$$\min_{\kappa\in(0,1)} \frac{\mu\gamma}{2(1-\gamma)\kappa(1-\kappa)}$$

$$\text{s.t. } \kappa \geq 1-\gamma$$

The solution is $\kappa = 1/2$ when $\gamma > 1/2$ and $\kappa = 1-\gamma$ when $\gamma \leq 1/2$ yielding the lower bounds of the variance to be

$$\sigma^2 > \frac{\mu}{2(1-\gamma)^2} \quad \text{when } \gamma \leq \frac{1}{2} \quad \text{and } \sigma^2 > \frac{2\mu\gamma}{1-\gamma} \quad \text{when } \gamma > \frac{1}{2}.$$

"$(B) \Rightarrow (A)$"

According to Proposition 6 if $b(\theta) = \kappa\theta + d$ satisfies (17) and condition (8), then $b(\cdot)$ is the best action function of an SMFE; and since it is affine, it is also the best action function of an AAE. All that needs to be shown is that such $\kappa > 0$ and $d \in \mathbb{R}$ exist.

The fundamental is distributed according to

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta-\bar{\theta})^2}{2\sigma^2}\right).$$

So, $\frac{p'(\theta)}{p(\theta)} = -\frac{\theta-\bar{\theta}}{\sigma^2}$. This, along with $b'(\theta) = \kappa$ and $b''(\theta) = 0$ make equation (9), read:

$$\kappa\theta + d = \theta - \frac{\mu\gamma}{2(1-\gamma)\kappa}\frac{\theta-\bar{\theta}}{\sigma^2}.$$

Solving for κ and d, one obtains two solutions:

$$\kappa_+ = \frac{1}{2}\left(1+\sqrt{1-\frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right) \qquad d_+ = \frac{\mu\gamma}{2(1-\gamma)\sigma^2\kappa_+}\bar{\theta}$$

and

$$\kappa_- = \frac{1}{2}\left(1-\sqrt{1-\frac{2\mu\gamma}{(1-\gamma)\sigma^2}}\right) \qquad d_- = \frac{\mu\gamma}{2(1-\gamma)\sigma^2\kappa_-}\bar{\theta}.$$

For either of κ_+ or κ_- to be positive reals, it is needed that $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$.

The second requirement for $b(\theta) = \kappa\theta + d$ to qualify for an AAE best action function is that $\text{Var}(b) > \mu/2$ i.e. that $\kappa^2\sigma^2 > \mu/2$. This condition for the solution κ_+, d_+ implies that either $\sigma^2 > \frac{\mu}{1-\gamma}$ or $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$. Together with $\sigma^2 > \frac{2\mu\gamma}{1-\gamma}$, the restrictions require that

- either $\gamma \leq 1/2$ and $\sigma^2 > \frac{\mu}{2(1-\gamma)^2}$
- or $\gamma > 1/2$ and $\sigma^2 > \frac{2\gamma\mu}{1-\gamma}$.

These are exactly the conditions assumed in statement (B). So, the solution with slope κ_+ is always the best action function of an AAE. $\square$

C Distribution families for error amplification Analysis

C.1 Skew normal g

The skew normal distribution $SN(b_0, s, \chi)$ with parameters $b_0 \in \mathbb{R}$, $s \in (0, \infty)$, and $\chi \in \mathbb{R}$ (introduced by O'Hagan and Leonard 1976), is a continuous distribution over $\mathbb{R}$ with PDF

$$f_{SN}(b; b_0, s, \chi) = \frac{2}{s}\phi\left(\frac{b-b_0}{s}\right)\Phi\left(\chi\left(\frac{b-b_0}{s}\right)\right)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the PDF and cumulative distribution function (CDF) of the standard normal distribution ($N(0,1)$). The important variable here is χ that adds skewness to the distribution (notice that when $\chi = 0$, the distribution boils down to a Gaussian).

Using the fact that the skew normal distribution is conjugate for the normal distribution, if $b \sim SN(b_0, s, \chi)$, then $a_i \sim SN\left(b_0,\ \sqrt{s^2 - \mu/2},\ \chi s\left(s^2 - \mu(1+\chi^2)/2\right)^{-1/2}\right)$ (see Azzalini 1985). So, as long as information costs are not too large ($\mu < 2s^2/(1+\chi^2)$), the function R defined through (8) is the PDF of a skew normal probability distribution, and the players have a continuous best response to g.

Section 5.3 uses a subset of skew normal distributions, parameterized by the single variable χ. In particular, the family of distributions used is

$$SN(\chi) := SN(-\chi\sqrt{2/(\pi(1+\chi^2))}, 1, \chi),$$

so that $\mathbb{E}[b] = 0$.

C.2 Mixture normal g

Another parameterization of a familiy of g functions is an equal-weight mixture $MN(\beta, s)$ of two normal distributions that have the same variance s^2 and are centered at β and $-\beta$, respectively. The resulting PDF is

$$f_{MN}(b; \beta, s) = \frac{1}{\sqrt{2\pi}s}\left(\frac{1}{2}\exp\left(-\frac{(b-\beta)^2}{2s}\right) + \frac{1}{2}\exp\left(-\frac{(b+\beta)^2}{2s}\right)\right).$$

Clearly, when $\beta = 0$, the distribution is a Gaussian. The mixture normal distribution is conjugate for the normal distribution: when $b \sim MN(\beta, s)$, then $a_i \sim MN(\beta, \sqrt{s^2 - \mu/2})$.

As long as $s^2 > \mu/2$, the function R defined through (8) is the PDF of a mixture normal probability distribution, and the players have a continuous best response to g.

Section 5.3 uses a subset of mixture normal distributions, parameterized by the single variable β. In particular, the family of distributions used is

$$MN(\beta) := MN(\beta, 1).$$

D Omitted proofs and calculations

D.1 Proof of Lemma 2

D.1.1 Case 1: $\mu = 0$

As $\mu = 0$, player i can obtain full information on the value of θ without paying any costs. So, conditional on the value of θ, her optimization problem becomes

$$\max_{a_i} \quad -(1-\gamma)(a_i - \theta)^2 - \gamma(a_i - \bar{a}(\theta))^2$$

Taking a first order condition (the objective is concave in a_i), one obtains that the optimal action is given by

$$(1-\gamma)\theta + \gamma\bar{a}(\theta).$$

So, given any $\bar{a}(\cdot)$ and any value of θ, player i has a unique best action given by the expression $b(\theta) = (1-\gamma)\theta + \gamma\bar{a}(\theta)$. Thus, her best response is to assign a probability mass of 1 to that action (conditional on θ). That is, her best response is to use r_i given by $r_i(a_i|\theta) = \delta(a_i - b(\theta))$ with δ being Dirac's delta function (almost all θ).

D.1.2 Case 2: $\mu > 0$

For expositional clarity, in this proof, $A_i = \mathbb{R}$ denotes the action space, whereas $\Theta = \mathbb{R}$ denotes the state space. Consider variations of the strategy of player i. These variations will be of the type $\tilde{r} = r + \varepsilon\eta$ for some $\varepsilon > 0$. These variations should still be feasible. That is, for all θ, it is required that $r(\cdot|\theta) + \varepsilon\eta(\cdot|\theta)$ is a probability distribution over A_i. It is required, thus, that for all θ, $\int_{A_i} r(a_i|\theta) + \varepsilon\eta(a_i|\theta)\,\mathrm{d}a_i = 1$ which leads to the condition that for all θ, $\int_{A_i} \eta(a_i|\theta)\,\mathrm{d}a_i = 0$. It also has to be that $r(a_i|\theta) + \varepsilon\eta(a_i|\theta) \geq 0$

and so $\eta(a_i|\theta) \geq -r(a_i|\theta)/\varepsilon$ for all a_i and θ. From the above equations, the following is calculated:[24]

$$
\begin{aligned}
U(r_i+\varepsilon\eta, r_{-i}) &= -(1-\gamma)\int_\Theta\int_{A_i}(a_i-\theta)^2(r_i(a_i|\theta)+\varepsilon\eta(a_i|\theta))p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \\
&- \gamma\int_\Theta\int_{A_i}(a_i-\bar{a}(\theta))^2(r_i(a_i|\theta)+\varepsilon\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta. \qquad (19)
\end{aligned}
$$

Next, the following derivatives are calculated:

$$
\begin{aligned}
\left.\frac{\mathrm{d}U(r+\varepsilon\eta, r_{-i})}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} &= -(1-\gamma)\int_\Theta\int_{A_i}(a_i-\theta)^2\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \qquad (20)\\
&- \gamma\int_\Theta\int_{A_i}(a_i-\bar{a}(\theta))^2\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta
\end{aligned}
$$

and, denoting $H(a_i) := \int_\Theta \eta(a_i|\theta)p(\theta)\,\mathrm{d}\theta$,

$$
\begin{aligned}
\left.\frac{\mathrm{d}I(r+\varepsilon\eta)}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} &= \int_\Theta\int_{A_i}\log(r(a_i|\theta))\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta - \qquad (21)\\
&- \int_{A_i}\log(R_i(a_i))H(a_i)\,\mathrm{d}a_i.
\end{aligned}
$$

Since the variations considered have to be feasible, player i has to solve the following constrained optimization problem:

$$
\begin{aligned}
&\max_{r_i\in L^1(\Theta,p)} U(r_i, r_{-i}) - \mu I(r_i)\\
&\text{s.t.} \int_{A_i} r_i(a_i|\theta)\,\mathrm{d}a_i = 1 \qquad \text{for all } \theta\in\Theta.
\end{aligned}
$$

Using $k(\theta)$ to denote the Lagrange multiplier for the θ-constraint, the Lagrangian for player i's decision problem is

$$
\mathscr{L}(r_i, k(\theta)) = U(r_i, r_{-i}) - \mu I(\boldsymbol{\theta}, \boldsymbol{a}_i) - \int_\Theta k(\theta)\left(\int_{A_i} r(a_i|\theta)\,\mathrm{d}a_i - 1\right)p(\theta)\,\mathrm{d}\theta.
$$

Therefore, for any given $\theta\in\Theta$ and all possible perturbations η, an optimal strategy r should satisfy the following first order conditions:

$$
\left.\frac{\mathrm{d}\mathscr{L}(r_i+\varepsilon\eta, k(\theta))}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 0 \Rightarrow
$$

[24]The effect of the other players' strategies is incorporated in $\bar{a}(\theta)$.

$$\int_{\Theta}\int_{A_i}\left[u_i(a_i,\theta)-\mu\log\left(\frac{r(a_i|\theta)}{R_i(a_i)}\right)-k(\theta)\right]\eta(a_i|\theta)p(\theta)\,\mathrm{d}a_i\,\mathrm{d}\theta=0 \tag{22}$$

$$\text{and}\int_{A_i}r_i(a_i|\theta)\,\mathrm{d}a_i=1\qquad\text{for all }\theta\in\Theta. \tag{23}$$

Where

$$u_i(a_i,\theta)=-(1-\gamma)(a_i-\theta)^2-\gamma(a_i-\bar{a}(\theta))^2. \tag{24}$$

Since condition (22) has to be satisfied *for all* η, it has to be the case that

$$-(1-\gamma)(a_i-\theta)^2-\gamma(a_i-\bar{a}(\theta))^2-\mu\left[\log(r_i(a_i|\theta))-\log(R_i(a_i))\right]=k(\theta)\text{ for all }\theta\in\Theta.$$

So $r(a_i|\theta)$ has to be:

$$r(a_i|\theta)=R_i(a_i)\exp\left(-\frac{k(\theta)}{\mu}\right)\exp\left(\frac{u_i(a_i,\theta)}{\mu}\right) \tag{25}$$

and, denoting $K(\theta):=\exp\left(-\frac{k(\theta)}{\mu}\right)$, (25) can be rewritten as

$$r(a_i|\theta)=R_i(a_i)K(\theta)\exp\left(\frac{u_i(a_i,\theta)}{\mu}\right). \tag{26}$$

All that remains to be done is to determine the functions $K(\cdot)$ and $R_i(\cdot)$. Now, from the definition of $R_i(a_i)$:

$$R_i(a_i)=\int_{\theta}r(a_i|\theta)p(\theta)\,\mathrm{d}\theta\Rightarrow\int_{\Theta}\frac{r(a_i|\theta)}{R_i(a_i)}p(\theta)\,\mathrm{d}\theta=1.$$

After substituting from (26) and (24), simple calculations give

$$\int_{-\infty}^{+\infty}K(\theta)\exp\left(-\frac{(a_i-b(\theta))^2}{\mu}\right)\exp\left(-\frac{\gamma(1-\gamma)}{\mu}(\theta-\bar{a}(\theta))^2\right)p(\theta)\,\mathrm{d}\theta=1. \tag{27}$$

In the above, $b(\theta)=(1-\gamma)\theta+\gamma\bar{a}(\theta)$. By assumption (smooth, monotone, full-support strategy profile), b is invertible with b^{-1} being the inverse of b. Because of assumption 2 of Definition 5, $b(\cdot)$ is bijective and strictly increasing. Thus, $\lim_{b\to\infty}b^{-1}(b)=\infty$ and $\lim_{b\to-\infty}b^{-1}(b)=-\infty$. So, with a change of the variable of integration from θ to $b=b(\theta)$, and by defining $G(\cdot)$ as

$$G(b)=\frac{K(b^{-1}(b))\exp\left(-\frac{\gamma(1-\gamma)}{\mu}(b^{-1}(b)-\bar{a}(b^{-1}(b)))^2\right)p(b^{-1}(b))}{(1-\gamma)+\gamma\bar{a}'(b^{-1}(b))} \tag{28}$$

condition (27) can be rewritten as

$$\int_{-\infty}^{+\infty} G(b)\exp\left(-\frac{1}{\mu}(a_i-b)^2\right)\mathrm{d}b = 1. \tag{29}$$

Notice that the above condition has to hold for all $a_i \in \mathbb{R}$. This can only happen if $G(b) = 1/\sqrt{\pi\mu}$.

Proof. Notice that the left-hand side of equation (29) is the convolution of G and f given by $f(x) = \exp\bigl(-x^2/\mu\bigr)$. Now, take the Fourier transform on both sides and use the convolution theorem:

$$\begin{aligned}
\mathscr{F}_{a_i}[(G*f)(a_i)](\xi) &= \mathscr{F}_{a_i}[1](\xi) \Rightarrow \mathscr{F}_{a_i}[G(a_i)](\xi)\cdot\mathscr{F}_{a_i}[f(a_i)](\xi) = \delta(\xi) \\
\Rightarrow \mathscr{F}_{a_i}[G(a_i)](\xi) &= \frac{1}{\sqrt{\pi\mu}}\exp\bigl(\mu\pi^2\xi^2\bigr)\delta(\xi)
\end{aligned}$$

Where $\delta(\cdot)$ is Dirac's delta function. By taking the inverse Fourier transform on both sides, the statement is proven:

$$\begin{aligned}
G(b) &= \mathscr{F}_{\xi}^{-1}\left[\frac{1}{\sqrt{\pi\mu}}\exp\bigl(\mu\pi^2\xi^2\bigr)\delta(\xi)\right](b) \\
&= \frac{1}{\sqrt{\pi\mu}}\int_{-\infty}^{+\infty}\exp(2\pi\imath\xi x)\exp\bigl(\mu\pi^2\xi^2\bigr)\delta(\xi)\,\mathrm{d}\xi = \frac{1}{\sqrt{\pi\mu}}.
\end{aligned}$$

□

So now $K(\theta)$ can be calculated.

$$K(\theta) = \frac{1+\gamma(\bar{a}'(\theta)-1)}{p(\theta)\sqrt{\pi\mu}}\exp\left(\frac{\gamma(1-\gamma)}{\mu}(\theta-\bar{a}(\theta))^2\right) \tag{30}$$

Using (30) in (26) yields

$$r(a_i|\theta) = R_i(a_i)\frac{1+\gamma(\bar{a}'(\theta)-1)}{p(\theta)\sqrt{\pi\mu}}\exp\left(-\frac{(a_i-b(\theta))^2}{\mu}\right). \tag{31}$$

The solution has to also satisfy $\int_{-\infty}^{+\infty} r(a_i|\theta)\,\mathrm{d}a_i = 1$ for all θ. Again, changing the variable from θ to $b = b(\theta)$, this condition yields

$$\int_{-\infty}^{+\infty} R_i(a_i)\exp\left(-\frac{(b-a_i)^2}{\mu}\right)\mathrm{d}a_i = \sqrt{\pi\mu}\,p(b^{-1}(b))\bigl(b^{-1}\bigr)'(b). \tag{32}$$

Notice that the left-hand side of equation (32) is the convolution of R_i and f. Now, take the Fourier transform on both sides and use the convolution theorem

$$\begin{aligned} \mathscr{F}_{a_i}[R_i(a_i)](\xi)\cdot\mathscr{F}_b[f(b)](\xi) &= \sqrt{\pi\mu}\cdot\mathscr{F}_b[p(b^{-1}(b))\left(b^{-1}\right)'(b)](\xi) \Rightarrow \\ \mathscr{F}_{a_i}[R_i(a_i)](\xi) &= \exp\!\left(\mu\pi^2\xi^2\right)\cdot\mathscr{F}_b[p(b^{-1}(b))\left(b^{-1}\right)'(b)](\xi) \Rightarrow \quad (33) \\ R_i(a_i) &= \mathscr{F}_\xi^{-1}[\exp\!\left(\mu\pi^2\xi^2\right)\cdot\mathscr{F}_b[p(b^{-1}(b))\left(b^{-1}\right)'(b)](\xi)](a_i) \quad (34) \end{aligned}$$

If the expression above is the PDF of a probability distribution, then the solution is calculated by equation (26) which — after noticing that $g(b) = p(b^{-1}(b))\left(b^{-1}\right)'(b)$ is the PDF of the best action $b = b(\theta)$ — becomes

$$r_i(a_i|\theta) = R_i(a_i)\frac{b'(\theta)}{p(\theta)\sqrt{\pi\mu}}\exp\left(-\frac{(a_i - b(\theta))^2}{\mu}\right) \quad (35)$$

with

$$R_i(a_i) = \mathscr{F}_\xi^{-1}[\exp\!\left(\mu\pi^2\xi^2\right)\cdot\mathscr{F}_b[g(b)](\xi)](a_i). \quad (36)$$

This solution is *unique*. The analyticity of $p(\cdot)$ and $b(\cdot)$ (and, therefore, of $g(\cdot)$) ensures that the solution to the player's decision problem is actually in continuous strategies rather than in strategies that put positive probability mass on a countable set of actions i.e. discrete or strategies with both a discrete and a continuous part (see Matějka and Sims 2010, Proposition 2).

Now, for the "only if" part, if $R(\cdot)$ defined through (36) was not the PDF of a probability distribution, player i would not have a continuous best reply to r_{-i}. Because if she did, the marginal of her action would need to be defined by equation (36). □

D.2 Proof of Proposition 3

The conditions of the proposition on $P_{a_j|\theta}$ ensure that $\bar{a}(\theta)$ given by

$$\bar{a}(\theta) = \int_{j\in[0,1]\setminus\{i\}}\left(\int_{\mathbb{R}} a_j P_{a_j|\theta}(\mathrm{d}a_j|\theta)\right)\mathrm{d}j$$

is analytic in θ. Together with $\bar{a}'(\theta) > \frac{1-\gamma}{\gamma}$, this means that $b(\cdot)$, i.e., $\theta \mapsto \gamma\bar{a}(\theta) + (1-\gamma)\theta$, is an analytic and strictly increasing bijection. That is, $\left\{P_{a_j|\theta}\right\}_{j\neq i}$ is a smooth, monotone, full-support profile. As $b(\theta)$ is continuous in θ and P_θ is absolutely continuous, it follows that P_b is also absolutely continuous (with PDF $g(\cdot)$).

Use $\overline{M}$ and $\underline{M}$ to denote the sets

$$\overline{M} := \{\overline{\mu} \in \mathbb{R}_+ : \mu \in [0, \overline{\mu}] \Rightarrow \text{best response to } g(\cdot) \text{ is continuous}\}$$

$$\underline{M} := \left\{\underline{\mu} \in \mathbb{R}_+ : \mu \geq \underline{\mu} \Rightarrow \text{best response to } g(\cdot) \text{ is not continuous}\right\}.$$

Step 1: When $\mu = 0$, then player i's best response is given by $\delta(a_i - b(\theta))$ and $R_i(a_i) = \int_{-\infty}^{+\infty} \delta(a_i - b) g(b) \,\mathrm{d}b = g(a_i)$, i.e., P_{a_i} is absolutely continuous and player i's best response is continuous. So, player i's best response is continuous when $\mu = 0$. Therefore, $\overline{M}$ is nonempty, since $0 \in \overline{M}$.

Step 2: Assume that for some $\overline{\mu} > 0$ player i has a continuous best response to $g(\cdot)$, i.e., that $\overline{R}_i := \mathscr{F}_\xi^{-1}\left[\exp\left(\overline{\mu}\pi^2\xi^2\right)\hat{g}(\xi)\right]$ is a probability density function. Since $\overline{R}_i$ is a PDF, by Bochner's theorem (Bochner 1933; Rudin 1962, p.19), the function $\widehat{\overline{R}}_i$, given by $\widehat{\overline{R}}_i(\xi) = \exp\left(\overline{\mu}\pi^2\xi^2\right)\hat{g}(\xi)$, is positive definite. Consider some $\mu \in (0, \overline{\mu})$ and begin with the following observation:

$$\exp\left(\mu\pi^2\xi^2\right)\hat{g}(\xi) = \exp\left(-(\overline{\mu} - \mu)\pi^2\xi^2\right)\exp\left(\overline{\mu}\pi^2\xi^2\right)\hat{g}(\xi) = \exp\left(-(\overline{\mu} - \mu)\pi^2\xi^2\right)\widehat{\overline{R}}_i(\xi).$$

The expression $\exp\left(-(\overline{\mu} - \mu)\pi^2\xi^2\right)$ is the Fourier transform of the normal distribution $N(0, (\overline{\mu} - \mu)/2)$ and, thus, a positive definite function. So, $\exp\left(\mu\pi^2\xi^2\right)\hat{g}(\xi)$ is a positive definite function as the product of two positive definite functions. This means that $\mathscr{F}_\xi^{-1}\left[\exp\left(\mu\pi^2\xi^2\right)\hat{g}(\xi)\right]$ is the PDF of a probability distribution. Thus, player i has a continuous best response to $g(\cdot)$ when cost is μ and—since μ was arbitrarily chosen—for any $\mu \in [0, \overline{\mu}]$. Therefore, if the best response to $g(\cdot)$ is continuous when the cost is $\overline{\mu}$, then $\overline{\mu} \in \overline{M}$.

Step 3: Similarly, if player i's best response is not continuous for some $\underline{\mu} > 0$, her best response is not continuous for any $\mu \geq \underline{\mu}$. Therefore, if the best response to $g(\cdot)$ is not continuous when the cost is $\underline{\mu}$, then $\underline{\mu} \in \underline{M}$.

Step 4: When $\mu > 2\,\mathrm{Var}(\boldsymbol{b})$, the player's best response is not continuous, as that would imply a negative variance for R_i. So, the set $\underline{M}$ is nonemtpy.

Step 5: For any given μ the best response exists (Matějka and Sims 2010) and is either continuous or not continuous. This, together with the results of Steps 2 and 3, implies that the sets $\overline{M}$ and $\underline{M}$ partition $\mathbb{R}_+$.

Step 6: Using $\mu^* = \sup \overline{M} = \inf \underline{M}$ completes the proof. □

D.3 Proof of Proposition 4

Let $\tau_i(\cdot|b)$ be the PDF of player i's action a_i conditional on the best action being b. Then, from Bayes's rule, one gets:

$$\varrho_i(b|a_i) = \frac{\tau_i(a_i|b)g(b)}{R(a_i)} \tag{37}$$

As $b(\cdot)$ is bijective with inverse $b^{-1}(\cdot)$, one can derive τ_i from the result of Lemma 2 with a change of variable:

$$\begin{aligned}\tau_i(a_i|b) &= \frac{R_i(a_i)}{p(b^{-1}(b))\,(b^{-1})'(b)} \frac{1}{\sqrt{\pi\mu}} \exp\left(-\frac{(a_i-b)^2}{\mu}\right) \\ &= R_i(a_i)\frac{1}{g(b)}\frac{1}{\sqrt{\pi\mu}} \exp\left(-\frac{(a_i-b)^2}{\mu}\right)\end{aligned}$$

and comparing with (37), one obtains

$$\varrho_i(b|a_i) = \frac{1}{\sqrt{\pi\mu}} \exp\left(-\frac{(a_i-b)^2}{\mu}\right).$$

□

D.4 Statement of Lemma 8

The following are standard properties of the Fourier transform that are used in later proofs.

Lemma 8. *The mean of a random variable* $\boldsymbol{x}$ *with PDF* $p_{\boldsymbol{x}}$ *is given by*

$$\mathbb{E}(\boldsymbol{x}) = \frac{1}{-2\pi\imath}(\mathscr{F}_x[p_{\boldsymbol{x}}(x)])'(0) \tag{38}$$

and its variance is given by

$$\mathrm{Var}(\boldsymbol{x}) = \sigma_x^2 = \left(\frac{1}{-2\pi\imath}\right)^2 (\mathscr{F}_x[p_{\boldsymbol{x}}(x)])''(0) - (\mathbb{E}(\boldsymbol{x}))^2. \tag{39}$$

Proof. Follows from the definition of the Fourier transform. □

D.5 Statement and proof of Lemma 9

A key result for the characterization of continuous equilibria is Lemma 9.

Lemma 9. *Let $g(\cdot)$ be the distribution of the best action of a smooth, monotone, full-support strategy profile of player i's opponents that satisfies condition (8). In player i's best response, her expected action conditional on the best action being b is given by*

$$\alpha(b) := \mathbb{E}(\boldsymbol{a}_i|b) = b + \frac{\mu}{2}\frac{\mathrm{d}}{\mathrm{d}b}(\log(g(b))). \tag{40}$$

Proof.

To lighten notation, the player index i is suppressed in this proof. Denote by $\tau(a|b)$ the probability density of action a conditional on the best action being b in player i's best response. From Bayes's rule

$$\tau(a|b) = \frac{\varrho(b|a)R(a)}{g(b)}.$$

Using the property of the Fourier transform (see equation (38)), the expected action of player i, conditional on b, is

$$\alpha(b) = -\frac{1}{2\pi\imath}(\mathscr{F}_a[t(a|b)])'(0) = -\frac{1}{g(b)2\pi\imath}(\mathscr{F}_a[\varrho(b|a)R(a)])'(0)$$

and using the convolution theorem as well as the properties of the Fourier transform,

$$\alpha(b) = -\frac{1}{g(b)2\pi\imath}\left(\mathscr{F}_a[\varrho(b|a)] * (\mathscr{F}_a[R(a)])'\right)(0). \tag{41}$$

Now

$$\begin{aligned}
\mathscr{F}_a[\varrho(b|a)](x) &= \mathscr{F}_a\left[\frac{1}{\sqrt{\pi\mu}}\exp\left(-\frac{(a-b)^2}{\mu}\right)\right](x) \\
&= \frac{1}{\sqrt{\pi\mu}}\exp(-2\pi\imath bx)\mathscr{F}_a\left[\exp\left(-\frac{a^2}{\mu}\right)\right](x) \\
&= \exp(-2\pi\imath bx)\exp\left(-\mu\pi^2x^2\right) \equiv \psi(x)
\end{aligned} \tag{42}$$

and

$$(\mathscr{F}_a[R(a)])'(x) = \frac{\mathrm{d}}{\mathrm{d}\xi}\left(\exp\left(\mu\pi^2\xi^2\right)\cdot\mathscr{F}_{\tilde{b}}[g(\tilde{b})](\xi)\right)\Big|_{\xi=x}$$

$$= \underbrace{2\mu\pi^2 x \exp\left(\mu\pi^2 x^2\right) \mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](x)}_{\zeta_1(x)} + \underbrace{\exp\left(\mu\pi^2 x^2\right)\left(\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right]\right)'(x)}_{\zeta_2(x)} \tag{43}$$

So,

$$(\psi * \zeta_1)(0) = \int_{-\infty}^{+\infty} \zeta_1(y)\psi(-y)\,\mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} 2\mu\pi^2 y \exp\left(\mu\pi^2 y^2\right) \mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](y) \exp(2\pi\imath b y) \exp\left(-\mu\pi^2 y^2\right) \mathrm{d}y$$

$$= 2\mu\pi^2 \int_{-\infty}^{+\infty} \exp(2\pi\imath b y) y \mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](y)\,\mathrm{d}y = 2\mu\pi^2 \mathscr{F}^{-1}\left[y\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right](y)\right](b) = 2\mu\pi^2 \frac{1}{2\pi\imath} g'(b)$$

and

$$(\psi * \zeta_2)(0) = \int_{-\infty}^{+\infty} \zeta_2(y)\psi(-y)\,\mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} \exp\left(\mu\pi^2 y^2\right)\left(\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right]\right)'(y) \exp(2\pi\imath b y) \exp\left(-\mu\pi^2 y^2\right) \mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} \exp(2\pi\imath b y)\left(\mathscr{F}_{\tilde{b}}\left[g\left(\tilde{b}\right)\right]\right)'(y)\,\mathrm{d}y = \frac{2\pi}{\imath} b g(b)$$

Bringing everything together

$$\alpha(b) = -\frac{1}{g(b)2\pi\imath}\left((\psi * \zeta_1)(0) + (\psi * \zeta_2)(0)\right)$$

and, finally,

$$\alpha(b) = b + \frac{\mu}{2}\frac{g'(b)}{g(b)}.$$

□

D.6 Proof of Proposition 5

The proof is given in the following steps:

1. The variance of player i's action in her best response is given by

$$\mathrm{Var}(\boldsymbol{a}_i) = \mathrm{Var}(\boldsymbol{b}) - \mu/2 \tag{44}$$

2. and its variance conditional on the best action being b is given by

$$\mathrm{Var}(\boldsymbol{a}_i|b) = \frac{\mu}{2} + \frac{\mu^2}{4}\frac{\mathrm{d}^2}{\mathrm{d}b^2}(\log(g(b))). \tag{45}$$

3. Moreover, $b \mapsto \mathbb{E}(a_i|b)$ and $\theta \mapsto \mathbb{E}(a_i|\theta)$ are increasing functions.

Proof of Item 1:

Recall the property of the Fourier transform (Lemma 8):

$$\mathbb{E}(\boldsymbol{x}) = \frac{1}{-2\pi\imath}(\mathscr{F}_x[p_{\boldsymbol{x}}(x)])'(0)$$

Now, start from eq. (36), take the first derivative on both sides at $\xi = 0$ and multiply by $(-2\pi\imath)^{-1}$ to get $\mathbb{E}(\boldsymbol{a}_i) = \mathbb{E}(\boldsymbol{b})$. Proceed to take the second derivative on both sides of equation (36) at $\xi = 0$, multiply by $(-2\pi\imath)^{-2}$ and take into account that $\mathbb{E}(\boldsymbol{a}_i) = \mathbb{E}(\boldsymbol{b})$. Now, use the second part of Lemma 8 to get:

$$\mathrm{Var}(\boldsymbol{a}_i) = -\frac{\mu}{2} + \sigma_b^2. \tag{46}$$

□

Proof of Item 2:

Following the same process for $\tau(\cdot|b)$, yields

$$\mathrm{Var}(\boldsymbol{a}_i|b) = \left(\frac{1}{-2\pi\imath}\right)^2 (\mathscr{F}_a[\tau(a|b)])''(0) - (\mathbb{E}(\boldsymbol{a}_i|b))^2. \tag{47}$$

Now

$$(\mathscr{F}_a[\tau(a|b)])''(x) = \left(\mathscr{F}_a\left[\frac{\varrho(b|a)R(a)}{g(b)}\right]\right)''(x) = \frac{1}{g(b)}(\mathscr{F}_a[\varrho(b|a)R(a)])''(x)$$

and

$$(\mathscr{F}_a[\varrho(b|a)R(a)])''(x) = \left(\mathscr{F}_a[\varrho(b|a)] * (\mathscr{F}_a[R(a)])''\right)(x). \tag{48}$$

Taking (43) and calculating the derivative, one gets

$$\begin{aligned}(\mathscr{F}_a[R(a)])''(x) = \hat{R}''(x) = \\ \underbrace{4\mu^2\pi^4x^2\exp\left(\mu\pi^2x^2\right)\hat{g}(x)}_{\zeta_3(x)} + \underbrace{4\mu\pi^2x\exp\left(\mu\pi^2x^2\right)\hat{g}'(x)}_{\zeta_4(x)} \\ + \underbrace{2\mu\pi^2\exp\left(\mu\pi^2x^2\right)\hat{g}(x)}_{\zeta_5(x)} + \underbrace{\exp\left(\mu\pi^2x^2\right)\hat{g}''(x)}_{\zeta_6(x)}.\end{aligned}$$

Moreover, from (42)

$$\mathscr{F}_a[\varrho(b|a)](x) = \exp(-2\pi\imath bx)\exp\left(-\mu\pi^2x^2\right) \equiv \psi(x)$$

and

$$\begin{aligned}
(\psi * \zeta_3)(0) &= \int_{-\infty}^{+\infty} \zeta_3(y)\psi(-y)\,\mathrm{d}y = (2\mu\pi^2)^2\mathscr{F}_y^{-1}\left[y^2\hat{g}(y)\right](b) = -\mu^2\pi^2g''(b) \\
(\psi * \zeta_4)(0) &= \int_{-\infty}^{+\infty} \zeta_4(y)\psi(-y)\,\mathrm{d}y = 4\mu\pi^2\mathscr{F}_y^{-1}[y\hat{g}'(y)](b) = -4\mu\pi^2\left(g(b) + bg'(b)\right) \\
(\psi * \zeta_5)(0) &= \int_{-\infty}^{+\infty} \zeta_5(y)\psi(-y)\,\mathrm{d}y = 2\mu\pi^2g(b) \\
(\psi * \zeta_6)(0) &= \int_{-\infty}^{+\infty} \zeta_6(y)\psi(-y)\,\mathrm{d}y = \mathscr{F}_y^{-1}[\hat{g}''(y)](b) = -4\pi^2b^2g(b)
\end{aligned}$$

Substituting the above together with $\mathbb{E}(\boldsymbol{a}_i|b) = b + \frac{\mu}{2}\frac{g'(b)}{g(b)}$ into (47) yields the result:

$$\mathrm{Var}(\boldsymbol{a}_i|b) = \frac{\mu}{2} + \frac{\mu^2}{4}\frac{\mathrm{d}^2}{\mathrm{d}b^2}\log(g(b))$$

□

Proof of Item 3:

From equation (40), one gets:

$$\alpha'(b) = 1 + \frac{\mu}{2}\frac{\mathrm{d}^2}{\mathrm{d}b^2}\log(g(b))$$

and, using the result of Item 2,

$$\mathrm{Var}(\boldsymbol{a}_i|b) = \frac{\mu}{2}\alpha'(b).$$

Now, since $\tau(\cdot|b)$ is a probability distribution, its conditional variance should be non-negative and, since g is analytic, well-defined (finite). So, since $\mathrm{Var}(\boldsymbol{a}_i|b) \geq 0$, the above equation leads to $\alpha'(b) \geq 0$. As $b'(\theta) > 0$ for all θ, $\mathbb{E}(a_i|\theta)$ is an increasing function of θ. □

D.7 Proof of Proposition 6

Start with the following Lemma.

Lemma 10. *Consider a beauty contest with flexible information acquisition. Then all SMFE are essentially symmetric i.e. in equilibrium all players use strategies that are equal to the same strategy $\tilde{r}$ almost everywhere.*

Proof. As there is a continuum of players, any single player i cannot influence the average action taken by the population for any value of θ. This means that all players face the same decision problem. Recall that each player has a unique best reply (up to deviations of measure zero, see Lemma 2) to a smooth, monotone, full-support profile. Thus, in equilibrium, the strategies that the players are using should be equal to the same strategy $\tilde{r}$ almost everywhere. □

The proof of Proposition 6 follows.

"(A) ⇒ (B)"

In light of Lemma 10, since all players have essentially the same best response to the equilibrium profile, the average action of the population conditional on b is given by

$$\alpha(b) = b + \frac{\mu}{2}\frac{g'(b)}{g(b)}.$$

In equilibrium, the best action b should be the one that is generated by aggregating the best responses of the players, i.e.,

$$b = \gamma\alpha(b) + (1-\gamma)\theta(b)$$

and, therefore, in equilibrium

$$\theta(b) = b - \frac{\gamma\mu}{2(1-\gamma)}\frac{g'(b)}{g(b)}.$$

Moreover, $g(\cdot)$ should be the distribution that is generated by $\theta(\cdot)$, i.e., (see eq. (7))

$$g(b) = p(\theta(b))\theta'(b).$$

"(B) ⇒ (A)"

Firstly, if $\theta(\cdot)$ is the inverse of the best action function, then the best action's distribution has the PDF $g(b) = p(\theta(b))\theta'(b)$. Since $\mathscr{F}_{\xi}^{-1}[\exp(\mu\pi^2\xi^2)\hat{g}(\xi)]$ is a probability distribution, the unique best response to g is continuous (see Lemma 2).

The fact that $\theta(\cdot)$ and $g(\cdot)$ satisfy (9) says that the profile where all players best respond to $\theta(\cdot)$ (equivalently, $b(\cdot)$) gives rise to $\theta(\cdot)$ as the inverse of the best action function, i.e., that it is an SMFE. □

References

Angeletos, G.-M. and A. Pavan (2007). "Efficient Use of Information and Social Value of Information". *Econometrica* 75 (4), pp. 1103–1142.

Azzalini, A. (1985). "A Class of Distributions Which Includes the Normal Ones". *Scandinavian Journal of Statistics* 12 (2), pp. 171–178.

Bergemann, D. and S. Morris (2016). "Bayes correlated equilibrium and the comparison of information structures in games". *Theoretical Economics* 11 (2), pp. 487–522.

Blackwell, D. (1951). "Comparison of Experiments". In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, Calif.: University of California Press, pp. 93–102.

Blackwell, D. (1953). "Equivalent Comparisons of Experiments". *The Annals of Mathematical Statistics* 24 (2), pp. 265–272.

Bochner, S. (1933). "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse". *Mathematische Annalen* 108 (1), pp. 378–410.

Carlsson, H. and E. van Damme (1993). "Global games and equilibrium selection". *Econometrica* 61, pp. 989–989.

Denti, T. (2020). "Unrestricted Information Acquisition". *mimeo*.

Dewan, T. and D. P. Myatt (2008). "The qualities of leadership: Direction, communication, and obfuscation". *American Political Science Review* 102 (03), pp. 351–368.

Echenique, F. (2002). "Comparative Statics by Adaptive Dynamics and the Correspondence Principle". *Econometrica* 70 (2), pp. 833–844.

Goryunov, M. and A. Rigos (2020). *Discontinuous and Continuous Stochastic Choice and Coordination in the Lab*. Working Paper 2020:17. Lund University, Department of Economics.

Grossman, S. J. and J. E. Stiglitz (1980). "On the Impossibility of Informationally Efficient Markets". *The American Economic Review* 70 (3), pp. 393–408.

Hébert, B. M. and J. La'O (2020). *Information Acquisition, Efficiency, and Non-Fundamental Volatility*. Working Paper 26771. National Bureau of Economic Research.

Hellwig, C., S. Kohls, and L. Veldkamp (2012). "Information Choice Technologies". *American Economic Review* 102 (3), pp. 35–40.

Hellwig, C. and L. Veldkamp (2009). “Knowing What Others Know: Coordination Motives in Information Acquisition”. *The Review of Economic Studies* 76 (1), pp. 223–251.

Hellwig, M. F. (1980). “On the aggregation of information in competitive markets”. *Journal of economic theory* 22 (3), pp. 477–498.

Jung, J., J. H. J. Kim, F. Matějka, and C. A. Sims (2019). “Discrete Actions in Information-Constrained Decision Problems”. *The Review of Economic Studies* 86 (6), pp. 2643–2667.

Kozlovskaya, M. (2018). “Industrial espionage in duopoly games”. *Available at SSRN 3190093*.

Maćkowiak, B. and M. Wiederholt (2009). “Optimal sticky prices under rational inattention”. *American Economic Review* 99 (3), pp. 769–803.

Manzano, C. and X. Vives (2011). “Public and private learning from prices, strategic substitutability and complementarity, and equilibrium multiplicity”. *Journal of Mathematical Economics* 47 (3). Mathematical Economics II : Special Issue in honour of Andreu Mas-Colell, pp. 346–369.

Matějka, F. and A. McKay (2015). “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model”. *American Economic Review* 105 (1), pp. 272–98.

Matějka, F. and C. A. Sims (2010). “Discrete actions in information-constrained tracking problems”. *Princeton University manuscript*.

Milgrom, P. and J. Roberts (1994). “Comparing Equilibria”. *The American Economic Review* 84 (3), pp. 441–459.

Morris, S. and H. S. Shin (2002). “Social Value of Public Information”. *The American Economic Review* 92 (5), pp. 1521–1534.

Morris, S. and M. Yang (forthcoming). “Coordination and Continuous Stochastic Choice”. *The Review of Economic Studies*.

Myatt, D. P. and C. Wallace (2012). “Endogenous Information Acquisition in Coordination Games”. *The Review of Economic Studies* 79 (1), pp. 340–374.

O’Hagan, A. and T. Leonard (1976). “Bayes estimation subject to uncertainty about parameter constraints”. *Biometrika* 63 (1), pp. 201–203.

Rudin, W. (1962). *Fourier analysis on groups*. Vol. 121967. Wiley Online Library.

Sims, C. A. (2003). “Implications of rational inattention”. *Journal of Monetary Economics* 50 (3), pp. 665–690.

Vives, X. (1990). “Nash equilibrium with strategic complementarities”. *Journal of Mathematical Economics* 19 (3), pp. 305–321.

Vives, X. (2005). “Complementarities and Games: New Developments”. *Journal of Economic Literature* 43 (2), pp. 437–479.

Vives, X. (2008). *Information and Learning in Markets: The Impact of Market Microstructure*. New Jersey: Princeton University Press.

Yang, M. (2015). “Coordination with flexible information acquisition”. *Journal of Economic Theory* 158, Part B, pp. 721–738. Symposium on Information, Coordination, and Market Frictions.